IMPROVING ONLINE DECISION MAKING PROCESS BASED ON THE RANKING OF USER REVIEWS AND PRODUCT FEATURES

AZRA SHAMIM

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR
2015

ABSTRACT

The ubiquity of Web2.0 with the proliferation of blogs and social networks transformed the way people express their opinions about different entities, such as products and services. Online reviews have become a powerful source of information for customers and business that gauge customers' purchase intentions and enterprise strategies. The amount of user generated content has grown at a fast pace that forces users to gravitate through a number of online reviews in order to get decision oriented information, which is time consuming and tedious job. Consequently, a new line of research 'opinion mining' has emerged. Opinion mining techniques can help to alleviate the problem of information overload in online reviews by analyzing, summarizing and presenting peoples' opinions. Online reviews vary greatly in quality and it has become imperative to identify high quality reviews to enhance the decision making process. However, most of existing opinion mining techniques ignore the quality of reviews. Although some review quality evaluation approaches are discussed in the literature, however, the focus is not on users' preferences. Feature-based opining mining is required to provide a detailed feature-based summary in order to satisfy users' need. Different methods have been proposed in the literature which evaluate and rank product features. However, existing feature ranking methods utilized the overall user rating and semantic polarity to rank product features, and overlook opinion strength. In addition, the visualization of the opinion summary is orthogonal to review quality evaluation and feature ranking. Most of existing opinion visualizations present overall positive and negative semantic on each feature and are unable to reflect opinion-strength based summary. The objectives of this research work are to integrate high quality reviews and opinion strength in feature ranking and to present opinion-strength based summarization using a visualization technique. Existing factors

for review ranking have been investigated and significant factors were assimilated in the proposed methods according to the users' preferences. Similarly, current elements for feature ranking have been examined and were amalgamated with opinion strength in the proposed method. Seminars and an online web based questionnaire survey was conducted to get the users' inclinations about opinion visualization to propose an opinion-strength based visualization. A feature based opinion mining system was developed based on proposed methods and experimental results on real life data sets show that integration of review and feature ranking with strength-based feature level summary can improve the decision making process.

ABSTRAK

Kewujudan berterusan Web2.0 serta perkembangan pesat blog dan jaringan sosial telah mengubah cara masyarakat mengekspresikan pendapat mereka mengenai pelbagai entiti yang berbeza, seperti produk dan perkhidmatan. Ulasan dalam talian telah menjadi sumber maklumat yang sangat berpengaruh kepada pelanggan dan perniagaan untuk mengukur niat pembelian pelanggan serta strategi perusahaan. Jumlah kandungan yang dijana pengguna telah meningkat dengan pantas dan hal ini memberi desakan kepada pengguna untuk membaca banyak ulasan dalam talian (online) dalam usaha memperoleh maklumat berorientasikan keputusan, yang sebenarnya mengambil masa yang panjang serta membosankan. Hal ini telah membawa kepada kewujudan kaedah penyelidikan baru yang dikenali sebagai "opinion mining". Teknik "opinion mining" dapat membantu mengurangkan masalah maklumat berlebihan dalam ulasan dalam talian dengan menganalisis, merumus, dan menunjukkan pendapat pengguna. Kualiti ulasan-ulasan dalam talian saling berbeza dan hal ini telah membuatkan pengenalpastian ulasan yang bermutu tinggi untuk mempercepatkan proses pembuatan keputusan sangat penting. Walau bagaimanapun, kebanyakan teknik "opinion mining" yang sedia ada tidak menitikberatkan kualiti ulasan dalam talian. Meskipun terdapat beberapa pendekatan dalam penilaian kualiti ulasan yang dibincangkan dalam hasil kajian, tetapi perbincangan tersebut tidak berfokus kepada keutamaan pengguna. "Opinion mining" berasaskan ciri diperlukan untuk memberikan rumusan berasaskan ciri yang terperinci agar dapat memenuhi keperluan pengguna. Pelbagai kaedah berbeza telah dicadangkan dalam hasil kajian yang menilai dan menentukan ranking ciri-ciri produk. Namun begitu, kaedah pemeringkatan yang sedia ada menggunakan kadaran pengguna (user rating) dan polariti semantik (semantic polarity) secara keseluruhan untuk menentukan ranking ciri produk,

dan mengetepikan kekuatan pendapat. Selain itu, pengvisualan rumusan mengenai pendapat ialah ortogonal dalam mengkaji penilaian kualiti dan ranking ciri. Kebanyakan pengvisualan pendapat yang sedia ada memaparkan semantik positif dan negatif bagi setiap ciri dan tidak mampu untuk menunjukkan rumusan berasaskan kekuatan pendapat. Objektif kajian ini adalah untuk mengintegrasikan ulasan berkualiti tinggi dengan kekuatan pendapat dalam ranking ciri serta untuk memaparkan rumusan berasaskan kekuatan pendapat dengan menggunakan teknik pengvisualan. Faktor ranking ulasan sedia ada telah diselidik dan faktor yang signifikan telah diasimilasikan dalam kaedah yang dicadangkan berdasarkan keutamaan pengguna. Begitu juga, elemen semasa untuk ranking ciri telah dikaji dan digabungkan dengan kekuatan pendapat dalam kaedah yang dicadangkan. Seminar dan soal selidik dalam talian telah dijalankan untuk mengetahui kecenderungan pengguna mengenai pengvisualan pendapat untuk mencadangkan pengvisualan berasaskan kekuatan pendapat. Sistem “opinion mining” berasaskan ciri telah dibangunkan berdasarkan kaedah yang dikemukakan dan keputusan eksperimen terhadap set data hidup nyata menunjukkan bahawa integrasi ranking ulasan dengan ranking ciri dan rumusan mengenai tahap ciri berasaskan kekuatan pendapat mampu menambah baik proses pembuatan keputusan.

ACKNOWLEDGEMENT

I am first and foremost humbly grateful to Almighty Allah for the supreme blessing with which the learning I under took in mission of knowledge met of fruitful end. I am grateful to my supervisor; Dr. Vimala Balakrishnan for her valuable supervision, guidance, sympathetic attitude and encouragement throughout my research work. I would like to express my gratitude and appreciation to my family for their endless love and support during my life. Without their moral support, this dissertation would never have been completed. I am really thankful to Mr. Muhammad Ahsan Qureshi, Mr. Ali Shahbaz, Ms. Fozia Anwar, Ms. Bushra Fayyaz and Mr. Zeeshan Shah for their timely and unconditional help, which can never be paid back. I am highly indebted to all my friends and colleagues for their cooperation and assistance during my research work.

DEDICATION

To my loving parents and family

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Background

Currently, businesses spend a lot of money on focus groups and questionnaire surveys to determine customers' opinions, sentiments and experiences about their products and services in the form of structured studies (Moghaddam & Ester, 2013). However, problems with these structured studies are cost, limitations imposed on free expression, design, administration and the missing opinions of a whole segment of the population (Kongthon, Haruechaiyasak, Sangkeettrakarn, Palingoon, & Wunnasri, 2011). With the increased use of the Web and the Internet, customers express their opinions and experiences via blogs, newsgroups, discussion boards and through writing reviews on websites (Na, Thet, & Khoo, 2010). As a result, a large amount of user generated data have been transferred to different online platforms (Li, Liao, & Lai, 2012), and is growing rapidly (Keikha & Crestani, 2010). Consequently, the Web consists of huge volumes of publicly available opinion data. This less structured 'word-of-mouth' (WOM) decision-oriented resource provides an alternative opportunity over focus groups and questionnaires to gather customers' feedback.

The development of Web 2.0 with rapid growth of social media has shifted the content publishing from businesses towards customers (Brien, 2011). Social media provide social interaction, using highly accessible and scalable communication techniques to create and exchange user generated content (Kaplan & Haenlein, 2010). This user generated content establishes a rich source of freely available opinion data, that is valuable to different stakeholders, such as enterprises, customers and service managers, with diverse information needs (Hao et al., 2013; Rohrdantz, Hao, Dayal, Haug, & Keim, 2012a). Moreover, social media have raised the level of sophistication of online shoppers, hence customers compare competing brands of products before making a purchase (Dalal & Zaveri, 2014). The opinionated postings in social media reshaped businesses and swayed public sentiments and emotions (Buche, Chandak, & Zadgaonkar, 2013).

Online reviews composed by many independent reviewers have become a powerful source of information for customers and businesses that significantly gauge customers' shopping behavior and enterprise strategies (Lipsman, 2007; Vermeulen & Seegers, 2009). Electronic word-of-mouth (e-WOM) significantly influences other customers' purchase intentions, product choice, the adoption and use of products and services (Jalilvand & Samiei, 2012). Houser and Wooders (2006) showed that positive user generated content has a significant impact on customers' decision-making process. It can also help to improve customers' satisfaction, build customers' trust and loyalty over time (Wu, Wei, Liu, & Au, 2010). Literature supports that positive user generated content has a positive correlation with the sale of a product (Chevalier & Mayzlin, 2006), whereas online customer complaints can easily reduce customers' loyalty and patronage, and create negative word-of-mouth (Buhalis, 2009).

The explosive growth of online opinion platforms, i.e. blogs, forum discussions, consumer feedback from emails and tweets provide another opportunity to entrepreneurs over focus groups, questionnaires, opinion polls and consultants for obtaining customers' reviews freely. Although there are numerous sources of user generated content, however, none of them is as focused as online reviews (Moghaddam, 2013). As a result, customers and entrepreneurs are increasingly using online product reviews for their purchase decisions and business planning, respectively (Zhang, 2012). Enterprises are now analyzing customers' online reviews from different online sources, such as Amazon, Rateitall, Cnet, Epinions, and TripAdvisor to assist their business decision-making process (Moghaddam, 2013).

The analysis of online reviews supports entrepreneurs in many business-intelligence tasks. It highlights the relative strengths and weaknesses of products, enterprise risks and threats from competitors (Liu, 2012; Xu, Liao, Li, & Song, 2011). Risk management, market intelligence, new product design and advertisement placement (i.e. placing an ad when one praises a product and placing another from a competitor if one criticizes a product) are also assisted by this analysis (Ganesan & Kim, 2008; Maynard, 2013; Xu et al., 2011). The prediction of future sales is also mined by the analysis of online reviews (Liu, 2012). Further, it sets a benchmark for products and services (Moghaddam & Ester, 2012). From the customers' point of views, the analysis supports a customer to identify the strengths and weaknesses of products for making a purchase decision and assists them in product search and comparison (Liu, 2012; Moghaddam & Ester, 2012).

The amount of user generated content has grown at a fast pace as the ubiquity of the Web has enabled easy participation of all Internet users through blogs, forums, wikis, twitter messages, companies' online surveys, feedback forms, news feeds and online news websites among others (Rohrdantz, Hao, Dayal, Haug, & Keim, 2012b). However, the growing volume of online reviews forces users to gravitate through a number of online reviews in order to get decision-oriented information, which can be time consuming and tedious (Ghose & Ipeirotis, 2007). Moreover, due to cognitive and physical limitations, people face difficulties in producing consistent results when the amount of opinion information to be analyzed is massive (Zhang, 2012). Therefore, there is a growing need to analyze and summarize a large collection of reviews automatically to overcome subjective biases and mental limitations by developing automated opinion mining systems (Wu et al., 2010; Zhang, 2012). Opinion mining techniques can help to alleviate the problem of information overload in online reviews by analyzing, summarizing and presenting people's opinions. Consequently, a new line of research 'opinion mining' has emerged to analyze people's opinions and sentiments from user generated content (Liu, 2012).

The organization of rest of the chapter is as follows. The general idea about the opinion mining field is presented in Section 1.2 followed by the key concepts about the area which are presented in Section 1.3. Problem statement is defined in Section 1.4 and the aim of the research is described in Section 1.5. Research objectives and research questions are highlighted in Section 1.6. Section 1.7 and 1.8 state research contributions and research significance. Methodology used in this research work is discussed in Section 1.9 and thesis outline is presented in Section 1.10.

1.2 Opinion Mining

Opinion mining is also known as sentiment mining, semantic analysis, opinion extraction and sentiment extraction. It is a recent discipline at the crossroads of Information Retrieval and Computational Linguistics, which tries to detect the opinions expressed in the natural language texts automatically (Cheng & Xu, 2008). Opinion mining is concerned not with the topic a document is about, but with the opinion it expresses. It primarily focuses on opinions, which express or imply positive or negative sentiments (Liu, 2012). It studies the extraction of opinions or sentiments from a given piece of text using methods from Text Mining, Natural Language Processing, Information Retrieval, Machine Learning, Web Data Mining and Computational Linguistics (Maynard, 2013). More formally, it analyzes people's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities, such as products, services, organizations, individuals, issues, events, topics and their attributes (Liu, 2012). The objectives of opinion mining include mining, summarizing, and visualizing people's opinions about different entities from online reviews. Specifically, opinion mining is the area of research that attempts to develop automatic systems to extract opinions from a text written in natural language (El-Halees, 2013; Liu, 2012).

Definition (Opinion Mining): Given a set of evaluative text documents *D* that contains opinions or sentiments about an object *O*, opinion mining aims to extract attributes and components of the object that have been commented on in each document $d \in D$ and to determine whether the comments are positive, negative or neutral (Liu, 2012).

An opinion mining system aims to generate a list of a product's significant features, determines the positive and negative comments on each feature, and finally produces a structured opinion summary. Opinion mining has become a popular research topic due to its widespread range of applications, such as news (Gamon et al., 2008; Koppel & Shtrimberg, 2004; Wanner, Rohrdantz, Mansmann, Oelke, & Keim, 2009), movie reviews (Pang, Lee, & Vaithyanathan, 2002; Zhuang, Jing, & Zhu, 2006), education (Binali, Potdar, & Wu, 2009), citation analysis (Piao, Ananiadou, Tsuruoka, Sasaki, & Mcnaught, 2007), government intelligence (Stylios et al., 2010), and product reviews (Funk, Li, Saggion, Bontcheva, & Leibold, 2008; Saggion, Funk, Street, & Sheffield, 2009) among others.

1.3 Basic Components of an Opinion

This section presents the basic components of an opinion and the key concepts of opinion mining. There are three basic components of an opinion, namely, opinion holder, object and opinion as shown in Figure 1.1. These components are described below:

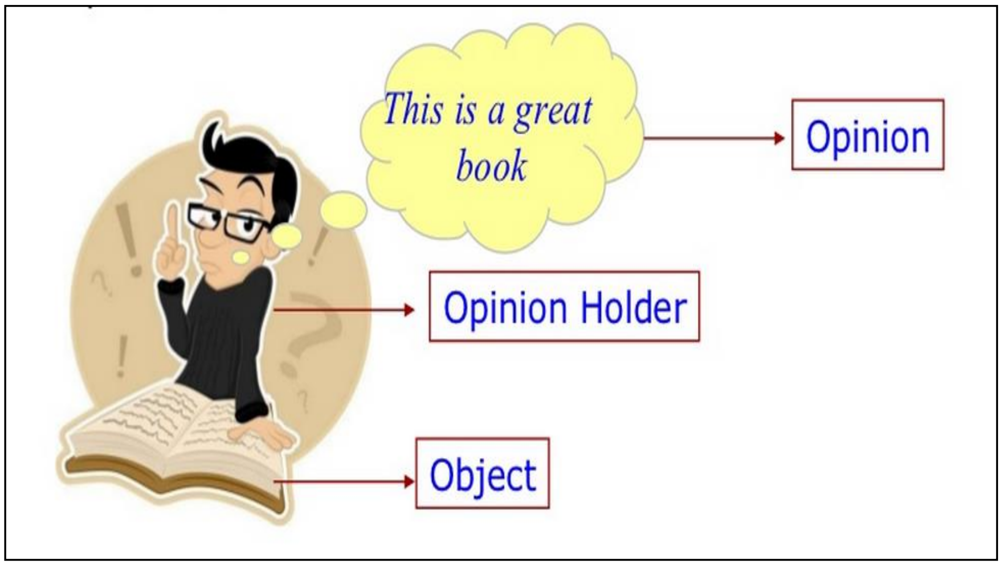


Figure 1.1: Components of an Opinion (Seerat & Azam, 2012)

1.3.1 Opinion Holder

The holder of an opinion is a person or organization that expresses a specific opinion on a particular object (Liu, 2006). In Figure 1.1, the person is the opinion holder. In the case of product reviews, opinion holders are usually the authors of the posts.

1.3.2 Object (Entity)

Opinions can be expressed on anything such as products, services, individuals, organizations, events, or topics, by any person or organization (Zhang, 2012). An object is a concrete or abstract item on which an opinion is expressed (Moghaddam & Ester, 2012). The general term object is used to denote the entity that has been commented on by opinion holders. In Figure 1.1, the book is the object on which the opinion holder (person) expressed the comment “This is a great book”.

Definition (Object): An object *O* is a product, service, person, event, organization, or topic. An object *O* is described by a pair, O: (T, W) (Liu, 2012), where

- *T* is a hierarchy of components, sub-components of object *O*, and so on.
- *W* is a set of attributes of object *O*
- Each node represents a component and is associated with a set of attributes of the component.
- *O* is the root node, which has a set of attributes.
- In this hierarchy, the root is the object *O* itself.
- Each non-root node is a component or sub- component of the object *O*.
- Each link is a part-of relationship.
- Each node is associated with a set of attributes.
- An opinion can be expressed on any node (component or sub-component) or attribute of the node.

Example: A particular camera model, *‘Canon Power Shot G3’* is an object. It has a set of attributes, i.e. picture quality, size and weight, and a set of components, i.e. lens, viewfinder and battery. The battery component also has its own set of attributes, i.e. battery life and battery weight.

An object *O* can be represented as a tree or hierarchy based on the definition as shown in Figure 1.2 for ‘*Canon Powershot G3*’ camera.

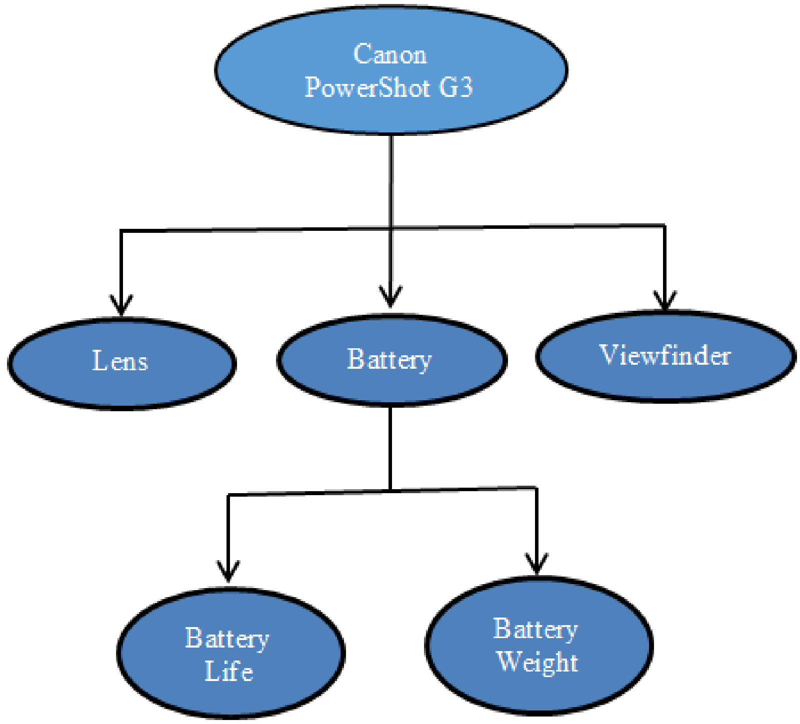


Figure 1.2: Components and Sub-components of Canon PowerShot G3

Reviewers can express an opinion on the root node (Canon PowerShot G3), e.g., '*I love Canon PowerShot G3*', or on any one of its attributes, e.g., '*The picture quality of Canon PowerShot G3 is excellent*'. Similarly, the components and sub-components of Canon PowerShot G3 can also be commented by reviewers, e.g., '*The battery life is short*'. The term feature can be used to represent both components and attributes for simplicity.

1.3.2.1 Feature

The term feature can be used interchangeably with aspect. A feature is an attribute of a product that is of interest to customers (Kunpeng, Narayanan, & Choudhary, 2010).

Definition (Feature): Feature is an attribute or component of an object *O* that has been commented on in an evaluation document *D* (Liu, 2006).

Definition (Feature Expression): A feature expression is an actual word or phrase that has appeared in reviews indicating a feature (Zhang, 2012).

Example: Picture, battery, size and weight are the features of the '*Canon PowerShot G3 camera*'. There are many feature expressions that can indicate the feature '*Picture*', e.g., '*photo*', '*pics*', and '*photographs*'.

Feature expressions are usually nouns and noun phrases, however, verbs, verb phrases, adjectives, and adverbs also indicate feature expressions in some cases. Liu (2006) showed that 60-70% of the features are nouns. Features can be classified as explicit and implicit based on their feature expressions. If a feature appears in a review sentence, it is called an explicit feature, otherwise it is an implicit feature (Hu & Liu, 2004). For instance, consider the following two sentences:

Sentence 1: '*The picture quality is good.*'

Sentence 2: '*The camera is too large.*'

'*picture quality*' in sentence one and '*camera size*' in sentence two are explicit and implicit features of a given camera, respectively. '*large'* is an implicit feature expression in the sentence two, which implies the feature '*size*'. Hence, in sentence two, *'size'* is an implicit feature.

Nouns and noun phrases in a sentence indicate explicit features, whereas other types of feature expressions encode implicit features. Many implicit features are adjectives and adverbs, e.g. expensive (price), and reliably (reliability).

Another category of features is the predefined features, which are known features provided by review websites such as picture quality, battery, so that users explicitly assign ratings to them.

1.3.3 Opinion

Sentiment is often used as synonyms of opinion, which refers to a semantic about an object or feature of a target object. An opinion is a subjective belief, and is the result of emotion or interpretation of facts (Moghaddam & Ester, 2012). More formally, it is a subjective statement, view, attitude, emotion, or appraisal about an object or feature of the object from an opinion holder (Liu, 2012). In Figure 1.1 (Section 1.3), 'This is *a great book*' is the opinion, which was commented by the opinion holder (person) on the object '*book*'.

Formally, an opinion can be described by two key components: a target *g* and a sentiment *s* on the target *g*, i.e., (g, s), where

- *g* can be any object or feature of the object about which an opinion has been expressed.
- *s* is a positive, negative, or neutral sentiment, or a numeric rating score (1–5 stars), which expresses the strength/intensity of the sentiment (Liu, 2012).

For instance, ‘*good*’ is a sentiment for the feature ‘*picture quality*’ in sentence one, whereas ‘*too large*’ is a sentiment for the feature ‘*camera size’* in sentence two based on the example in Section 1.3.2.1. Opinion words are used to describe the semantic on features or target objects. An opinion can be explicit or implicit.

Definition (Explicit Opinion): An opinion which is explicitly expressed on feature *f* in a sentence (Liu, 2012).

Definition (Implicit Opinion): An opinion on feature *f* implied in a sentence (Liu, 2012).

Example: Consider the following sentences:

Sentence 3: ‘*The picture quality of this phone is amazing.*’

Sentence 4: ‘*The headset broke in one day.*’

Sentence three expresses an explicit positive opinion on the feature ‘*picture quality*’ while sentence four depicts an implicit negative opinion on the feature ‘*headset*’.

Specifically, an opinion is a quintuple, (o_j, f_{jk}, s_{ijkl}, h_k, t_l) (Liu, 2012), where

- o_j is the name of a target object
- f_{jk} is a feature of object o_j
- s_{ijkl} is the sentiment value of the opinion given by the opinion holder h_k on the feature f_{jk} for the object o_j at time t_l. The sentiment value of S_{ijkl} is positive, negative, or neutral. Neutral opinions are ignored in the output as they are not usually useful.

- h_k is the opinion holder
- t_l is the time when the opinion was expressed by the opinion holder h_k.

Figure 1.3 shows an opinion quintuple where object is '*Canon PowerShot G3 camera*'. '*Picture Quality*' is a feature for the object '*Canon PowerShot G3*'. The semantic value of the feature '*Picture Quality*' is positive, as good is a positive opinion word in the literature (Liu, 2012). The '*reviewer 1*' is the opinion holder who has commented on the feature '*Picture Quality*' for object '*Canon PowerShot G3*' at time '*T1*'. As a result, the opinion quintuple is (G3, Picture Quality, Positive, Reviewer 1, T1) as shown in Figure 1.3.

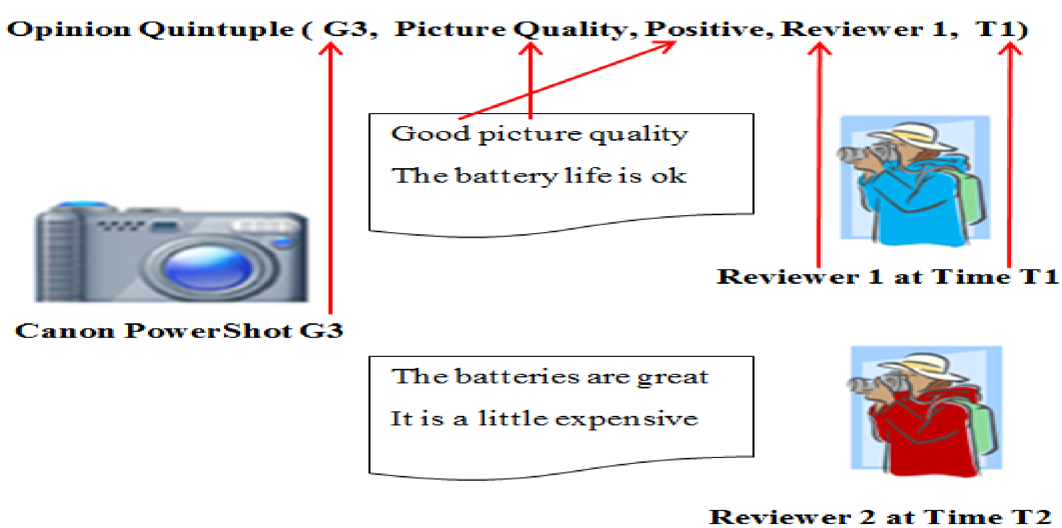


Figure 1.3: An example of an Opinion Quintuple

1.3.3.1 Semantic Orientation/Polarity

Semantic orientation refers to opinion orientation. It means whether the opinion on a feature or a target object is positive, negative or neutral (Ding, Liu, Yu, & Street, 2008).

Opinion words are commonly used to express positive or negative opinions. For instance, in sentence one (Section 1.3.2.1), the opinion word is '*good*', that describes a positive semantic orientation, while in sentence two, the opinion word is '*too large*', that presents a negative semantic orientation. Some of the common positive opinion words are amazing, good, great, fine, wonderful and lovely, whereas the common negative opinion words are poor, expensive, bad, and terrible.

1.3.3.2 Opinion Strength

Reviewers use different opinion words to describe target objects. Opinion words vary in term of opinion intensity they are expressing. Opinion strength measures the degree of polarity, positive or negative, in a subjective sentence (Raghavan, 2009). It reflects how positive or negative an opinion word is. It describes whether the positive opinion expressed by a text on a target object is Weakly Positive, Mildly Positive, or Strongly Positive (Binali et al., 2009; Osimo & Mureddu, 2012). For instance, the positive opinion word '*excellent*' is more positive than the positive opinion word '*good*'. Similarly, the negative opinions can be classified into Weakly Negative, Mildly Negative, or Strongly Negative (Binali et al., 2009; Osimo & Mureddu, 2012). The negative opinion word '*worst*' expresses more negative opinion on a target object than the negative opinion word '*bad*'. Consider the following sentences:

Sentence 5: '*The picture quality is excellent*'

Sentence 6: '*The picture quality is good*'

Sentence five and six is expressing positive opinions about the picture quality of two target products (cameras). However, if a customer wants to buy a camera based on these two opinions, he/she will prefer to purchase the first camera as sentence five is expressing more positive opinion as compared to sentence six.

1.4 Problem Statement

Online reviews have grown at a remarkable speed and vary greatly in quality, resulting in an information overload problem (Liu & Lin, 2007; Moghaddam, Jamali, & Ester, 2012a; Ngo-Ye & Sinha, 2012). Consequently, it becomes difficult to identify high quality helpful reviews to enhance the decision-making process. For this purpose, some review websites ask users to rate reviews and vote for their helpfulness. For instance, the reader of a review on amazon.com can indicate whether he/she finds a review helpful by responding to the question "*Was the review helpful to you?* " just below each review. The feedback results from all those responded are then aggregated and displayed right before each review, e.g., '*10 of 15 people found the following review helpful*'. Some websites display the percentage of positive and negative votes or the average rating. The helpfulness vote (how many people found the review helpful) and users' rating (ranges from 1 to 5 stars, where 1 star rating reflects an extremely negative view of product and 5 star rating indicates an extremely positive view of a product) project public endorsement, which may influence other customers' shopping behavior (Korfiatis, García-Bariocanal, & Sánchez-Alonso, 2012; Mudambi & Schuff, 2010; Walter, Battiston, & Schweitzer, 2011).

Commonly, users explicitly filter reviews based on their users' ratings (star ratings) and/or helpfulness votes in order to get high quality informative reviews, which cover a diverse set of opinions (Tsaparas, Ntoulas, & Terzi, 2011). Moreover, most of the existing opinion mining systems ignore the quality of reviews, therefore effective review quality evaluation methods are required to identify high quality reviews (Chen & Tseng, 2010). The quality of a review is a property orthogonal to its polarity or embedded opinions (Zhang & Varadarajan, 2006) or how helpful a review is (Kim, Pantel, Chklovski, & Pennacchiotti, 2006). Unhelpful or low quality reviews can be excluded from review summaries (Chen & Tseng, 2010). Some review quality evaluation approaches are discussed in the literature (Chen & Tseng, 2010; Ghose & Ipeirotis, 2007, 2011; Kim et al., 2006; Ngo-Ye & Sinha, 2012; Zhang & Varadarajan, 2006), however, the focus is not on the users' preferences that define the important parameters according to the users' perspectives.

Different methods have been proposed in the literature to evaluate and rank product features based on feature frequency, semantic orientation and users' rating (Eirinaki, Pisal, & Singh, 2012; Lei, Liu, Lim, & Eamonn, 2010; Li, Chen, & Tang, 2011; Moghaddam & Ester, 2010; Scaffidi et al., 2007; Yang, Kim, & Lee, 2010). However, current feature ranking methods based on users' ratings are not suitable to rank product's features, as users' rating projects entire product evaluation (Scaffidi et al., 2007; Yang et al., 2010). In other words, the rating portrays the product evaluation as a whole, but obscure individual features in terms of both positive and negative evaluation within the same review (Yang et al., 2010). Consider the following review with a 4 star rating:

'*The battery life is excellent. The flash is good. It provides outstanding ease of use. The LCD display is inaccurate. The self timer is bad.*'

In the above review, the user expressed positive opinions on the features; ‘*battery life*’, ‘*flash*’, ‘*ease of use*’. The user also provided negative evaluations on the features; ‘*LCD display*’ and ‘*self timer*’ in the same review, and a 4 star rating assigned to the target object (camera). Here, the 4 star rating does not mean that every feature mentioned in the review has been rated as 4 star, indicating that the use of the overall rating for feature ranking can be incorrect (Yang et al., 2010). Moreover, existing approaches overlook opinion strength, which defines how positive or negative an opinion word is, for instance, ‘excellent’ shows more positive semantic than ‘good’. Furthermore, existing feature ranking approaches also disregard the quality of reviews (Ahmad & Doja, 2012; Eirinaki et al., 2012; Lei et al., 2010; Scaffidi et al., 2007; Yang et al., 2010).

In addition, the visualization of the opinion summary along with review quality evaluation and feature ranking is equally important. A detailed feature-based summary with adequate visualization may be more useful than a summary that only shows an average score for product’s features (Yang et al., 2010). Recently, the topic of automatic opinion mining has been addressed (Pang & Lee, 2008), however, less efforts have been made for opinion visualization techniques. Therefore, not only automatic opinion mining algorithms for data analysis are imperative, but also opinion visualizations that appropriately convey hidden opinion trends to data analysts (Wu et al., 2010). Opinion visualizations have been shown to support the exploration of opinion data, however, the visualizations only present overall feature-based positive and negative evaluation and are unable to reflect opinion-strength-based summary (Liu et al., 2005; Oelke et al., 2009; Wanner et al., 2009; Wu et al., 2010).

The above discussion resulted in the following main problem statement:

"*Existing feature ranking methods and opinion visualization techniques neglect opinion-strength and quality of reviews according to users' preferences, resulting in imprecise and low-quality semantic summary*".

1.5 Aim of the Research

This thesis is devoted to integrating high quality reviews in feature ranking methods with opinion-strength-based visualization in order to provide customers with high quality decision-oriented information from enormous online reviews. For this purpose, in the first step, the problem of selecting high quality informative reviews according to users' preferences was addressed. In the next step, a new approach for feature ranking is proposed based on high quality reviews and opinion strength. In the last step, a visualization is introduced that allows a detailed insight into products' features and corresponding sentiments at different levels of opinion strengths. The primary aim of the research is to provide opinion-strength-based feature ranking and visualization using high quality reviews based on the users' preferences for the improvement of the decision-making process.

1.6 Research Objective and Questions

To achieve the aim, the research objectives and research questions of this work are discussed as follows:

Objective 1: To identify way(s) to incorporate users' preferences in ranking reviews

The number of online reviews is growing at a remarkable speed. Consequently, the quality of online reviews varies due to differences in the knowledge and experience of opinion holders. As a result, it is desirable to distinguish high quality reviews from low quality reviews in order to provide high quality decision-oriented information. In the literature, researchers proposed different review ranking methods, however, the focus in not on users' perspectives. The first objective is related to the integration of users' preferences in ranking reviews. The following two research questions are associated with this objective.

Research Question 1: What are the existing review ranking techniques?

Research Question 2: How to incorporate users' preferences in review ranking?

The first question is related to state-of-the-art review ranking approaches. To achieve objective one, first, knowledge of existing review quality frameworks is necessary to propose a new review ranking technique incorporating users' preferences. The second question is associated with the assimilation of users' preferences in review ranking. A literature survey is required to find way(s) to incorporate users' preferences in the proposed review ranking technique.

Objective 2: To enhance feature ranking using opinion-strength

The second objective is associated with feature ranking. A typical review provides both positive and negative evaluations on features of a target object and an overall users' rating for the target object. The feature ranking methods utilizing users' rating are incapable of presenting precise and high quality ranking, as users' rating reflects the product evaluation and indeterminate about individual feature evaluation in terms of both positive and negative evaluation within the same review. In order to address this issue and to achieve objective two, two research questions are defined.

> **Research Question 3: What are the existing feature ranking techniques?**
>
> **Research Question 4: How to enhance current feature ranking using opinion strength?**

First, a detailed review of existing feature ranking approaches is required, which will be addressed by answering research question three. Research question four addresses how to integrate opinion strength in feature ranking to achieve objective two.

Objective 3: To design an opinion-strength-based visualization based on users' preferences

The third objective of this work is related to the design and implementation of opinion visualization technique. Existing feature-based opinion visualization techniques reflect overall positive and negative evaluations or an average semantic on each feature of a target product, and are unable to portray different levels of opinion strengths. To address this problem, an opinion-strength-based opinion visualization technique is required, which incorporates users' preferences. First, a questionnaire survey is required to collect users' preferences about existing opinion visualization techniques. Then, based on the analysis of the collected data an opinion-strength-based opinion visualization technique will be proposed. The research question related to this objective is described below:

> **Research Question 5: How to present opinion-strength based summary using visualization techniques?**

Research question five addresses the ways in which opinion strength can be presented in feature-based opinion summary so that customers' and enterprises can investigate people's opinions at various levels of intensity.

Objective 4: To evaluate the effectiveness of the proposed review and feature ranking methods, and opinion visualization technique

The last objective of this work is related to the evaluation of proposed systems in term of its effectiveness. The following research question is associated with objective four.

> **Research Question 6: How the proposed system and opinion visualization technique can be evaluated to measure its effectiveness?**

Question six investigates different evaluation approaches to measure the effectiveness of the proposed review and feature ranking methods. Moreover, it explores ways to determine the usability and usefulness of the proposed opinion visualization technique.

1.7 Significant Research Contributions

The research contributions of this work are discussed below:

a) The first contribution of this work is related to the review ranking in which a method to rank the reviews according to their quality and users' preferences is developed. The proposed review ranking method is based on a superset of state-of-the-art review ranking features along with the features having relatively greater tendency to predict review helpfulness. This results in assimilation of the most influential factors, users' rating and helpfulness votes with the number of features and opinion words. The proposed review ranking method is dissimilar with existing studies as it integrates significant parameters of existing review ranking methods with important features that can predict the quality of reviews to a great extent.

b) The proposed method integrates users' preferences in review ranking, and thus is different from the state-of-art review ranking methods.

c) A method is developed to identify the relative importance of significant product features. Unlike existing studies, the opinions of the reviewers about the product and its features are considered in which the opinions are on a continuum from negative to positive, not simply the binary negative or positive. The use of opinion strength in feature ranking also results in more accurate feature ranking.

d) Another contribution is the introduction of opinion-strength-based opinion visualization that highlights critical product features and facilitates comparison between the positive and negative opinions of a particular feature with emphasis on opinion strength. In contrast to existing opinion visualization techniques, the proposed opinion-strength-based visualization technique provides an opinion summarization by which customers and enterprises can investigate people's opinion at various levels of intensity.

1.8 Research Significance

Traditional text processing techniques are often developed to glean factual information from natural language text. These techniques have been focused on retrieval and mining factual information, such as Web search, information retrieval, and many other text mining and natural language processing tasks. The development and overwhelming popularity of social media resulted in the generation of massive amounts of opinion data. This freely available opinion data significantly influence customers' buying behaviors and enterprise strategies. Traditional text processing techniques are unable to analyze opinion data as it is unstructured, ungrammatical, amorphous, noisy, difficult to deal with algorithmically, containing spelling errors (e.g. improper capitalization), abbreviations, slang and emoticons. This is the reason why the extraction of an opinion summary of opinion documents continues to be a challenging problem for opinion mining. The enormous size of the online reviews, the diversity of the comments, and the uneven distribution of feedback over time make sentiment analysis very challenging.

In this research work, the existing efforts in feature ranking, opinion summarization, and visualization are redirecting towards a novel perception by which customers' and enterprises can investigate people's opinion at various levels of intensity with high quality decision-oriented information.

1.9 Research Methodology

This study uses both quantitative and qualitative methods. The research methodology used in this thesis is shown in Figure 1.4. The study involves the following steps:

i. Conducting a literature review to identify factors that affect the quality of a review.

ii. Conducting a literature review to investigate factors that are currently used for feature ranking.

iii. Conducting a review of state-of-the-art opinion visualization techniques.

iv. Problem identification from the literature review.

v. Administering a questionnaire survey to collect the users' preferences about existing opinion visualization techniques.

vi. The development of a system based on the proposed review and feature ranking along with opinion-strength-based visualization.

vii. The evaluation of the proposed system in terms of accuracy.

viii. The evaluation of the proposed visualization in terms of usability.

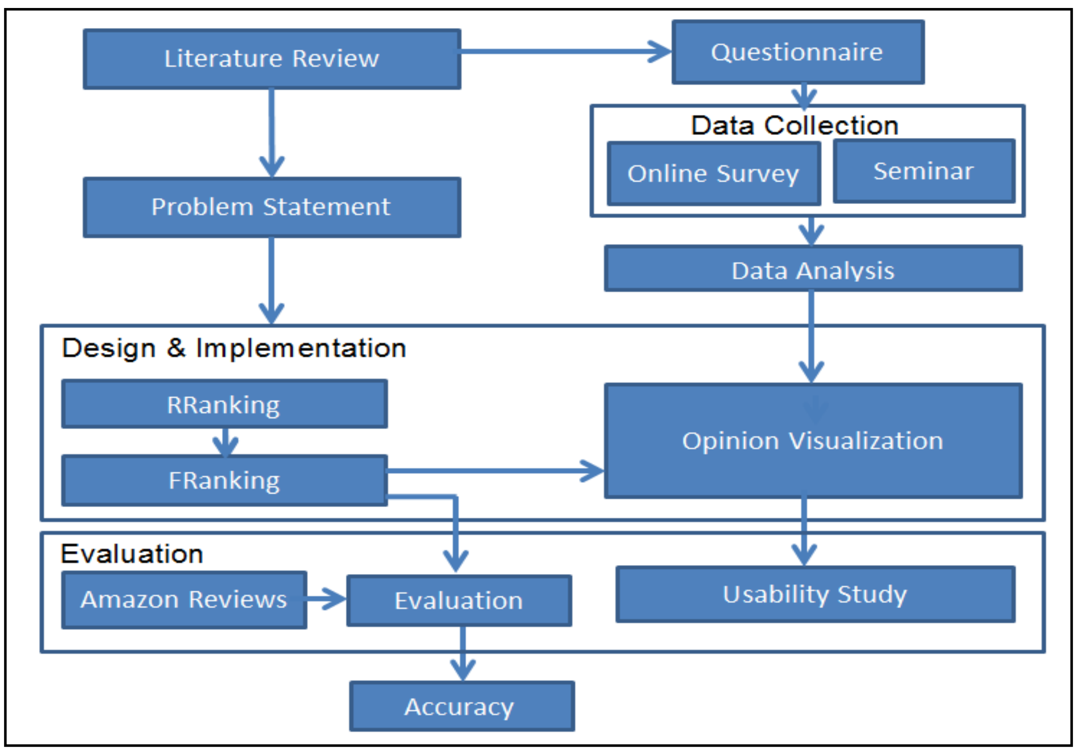


Figure 1.4: Research Methodology

1.10 Thesis Outline

This thesis consists of six chapters that include introduction, opinion mining, literature review, methodology, results and discussion, conclusion, limitation and future work. The detailed organization of rest of the thesis is as follows:

Chapter 2: Opinion Mining

This chapter discusses demands and potential applications of opinion mining, types of opinion, levels of semantic analysis, feature-based opinion mining, objectives, tasks and phases of opinion mining. Further, it provides an overview of evaluation measures and review formats.

Chapter 3: Literature Review

This chapter describes a comprehensive review of existing review quality evaluation approaches. It also presents state-of-the-art techniques for feature-based opinion mining, feature ranking and opinion visualization. Moreover, the research issues and challenges of opinion mining are highlighted in this chapter.

Chapter 4: Methodology

This chapter presents the prototype system which has been developed based on the proposed review and feature ranking methods. It also describes experimental data set and setup. Moreover, it discusses the methodology used for the evaluation of the proposed methods and opinion-strength-based opinion visualization.

Chapter 5: Results and Discussion

The chapter presents experimental results and discussion along with opinion-strength-based visualization. Moreover, the accuracy of the system is compared with a state-of-the-art system in this chapter.

Chapter 6: Conclusion, Limitation and Future Work

This chapter concludes this research work, and presents limitations and various perspectives for future research.

Chapter 2 : Opinion Mining

Chapter 1 introduced some basic concepts about opinion mining in Sections 1.2 and 1.3. However, detailed background knowledge about opinion mining is needed to understand the contributed research work of this thesis. Therefore, this chapter discusses the demands for opinion mining, applications, basic terminology, general opinion mining tasks, objective, phases, levels of analysis, evaluation measures and review formats.

2.1 Demands for Opinion Mining

The explosion of social media services, such as review sites, newsgroups, forum discussions, blogs, and discussion board have made it possible to access a large pool of peoples' experiences and opinions. Today, businesses consult online reviews to pinpoint (i) the relative strengths and weaknesses of their products, (ii) analyze threats from competitors and enterprise risks, (iii) support decision-making and risk management, and (iv) design new products and marketing strategies (Xu et al., 2011). On the other hand, customers refer to online reviews for making an informed decision about the purchase of a product (J. Lee, Park, & Han, 2011).

Pang and Lee (2008) reported interesting statistics of two surveys (N=2000 American adults) on this review revolution:

- Eighty one percent of Internet users (or 60% of Americans) have done online research on a product at least once.
- Twenty percent (15% of all Americans) do so on a typical day.

- Between 73% and 87% readers of online reviews report that reviews had a significant influence on their purchase.
- Consumers report being willing to pay from 20% to 99% more for a 5-star-rated item than a 4-star-rated item.
- Thirty two percent have provided a rating and 30% (including 18% of online senior citizens) have posted an online comment.

Similarly, Canada's largest Internet marketing company reported similar statistics showing demand for opinion mining (Moghaddam, 2013):

- Traffic to the top 10 review sites grew on average 158% in 2009.
- Ninety two percent of online consumers have more confidence in online information than they get from a salesclerk or other sources.
- Seventy percent consult reviews or ratings before purchasing.
- Ninety seven percent who made a purchase based on an online review, found the review to be accurate.
- Seventy percent who read reviews share them with friends, family, and colleagues thus amplifying their impact.
- Thirty four percent have turned to social media to share their feelings about a company, 26% users expressed dissatisfaction, and 23% users shared companies or products they like.

Another study showed that 51% of consumers use the Internet even before making a purchase in shops (Moghaddam & Ester, 2013). Horrigan (2008) highlighted that a majority of American Internet users had a positive experience during online product research, however, 58% users also reported that online information was missing, impossible to find, confusing, and/or overwhelming (Horrigan, 2008).

Opinion mining is not only valuable to customers in the decision-making process about the purchase of a product (Popescu & Etzioni, 2005), but also essential for businesses to understand customers' opinions on their products and services (Ding et al., 2008). The above statistics emphasize the need for opinion mining systems for both customers and enterprises, as these systems provide an excellent opportunity to support many business related tasks, such as sales management, reputation management etc.

2.2 Applications of Opinion Mining

Currently, opinion mining plays an important role in diverse domains, i.e. business intelligence (Pang & Lee, 2008), government intelligence (Stylios et al., 2010), news (Nagar & Hahsler, 2012; Wanner et al., 2009), recommender systems (Pang & Lee, 2008), question answering (Somasundaran, Wilson, Wiebe, & Stoyanov, 2007), citation analysis (Pang & Lee, 2008), shopping (Xu et al., 2011), education and entertainment (Binali et al., 2009).

The field of opinion mining is well-suited to business intelligence as enterprises consult online reviews to identify the strengths and weaknesses of their products in order to design new products (Pang & Lee, 2008). On the other hand, it supports many businesses-intelligence tasks, such as sales management, reputation management, public relation, trend prediction, decision-making, risk management, and marketing strategies, and analyzes threats from competitors and enterprise risks (Ganesan & Kim, 2008; Liu, 2012; Moghaddam & Ester, 2013; Xu et al., 2011). The most widespread application of opinion mining is a decision support for customers. It assists customers in making purchase decisions by providing competitive intelligence (Xu et al., 2011). Government intelligence is another application of opinion mining (Pang & Lee, 2008). It empowers governments to monitor people's opinions on public policies as public opinions matter a lot in the government decision-making. Similarly, governments can predict what the public is thinking about pending policy, law, and legislative proposals (Stylios et al., 2010). It also enables election candidates to identify their strengths and weaknesses, public support or opposition, and to re-define their policies in accordance with electorate opinions (Bansal, Cardie, & Lee, 2008).

Opinion mining also has potential application in news analysis. It analyzes the emotional contents in news and highlights similar or redundant news items (Wanner et al., 2009). It also pinpoints interesting trends and peculiarities in news items. On the other hand, users can find the most popular articles, articles most emotionally discussed, articles most cited by liberals and conservatives, for example, the article *'A muslim belongs in the cabinet'* is the most popular article with 15 and five liberal and conservative views, respectively (Gamon et al., 2008). Citation analysis is another area where opinion mining can prove useful. It assists to identify whether an author is citing a piece of work as supporting evidence, dismisses the cited work, or to track literary reputation (Pang & Lee, 2008).

Opinion mining can also be augmented to recommender systems to recommend items that receive a great deal of positive feedback, and not to recommend items that receive a lot of negative feedback (Pang & Lee, 2008). Opinion mining has potential relation with question answering, for instance, it is better to answer opinion-oriented questions by including the information about how positively or negatively an entity is viewed by other users (Somasundaran et al., 2007). In addition, users can also access positive and negative comments on recent releases, popular TV programs, and movies using opinion mining tools, that guides users about which movies or program to watch (Binali et al., 2009). Similarly, it also improves the education system based on the analysis of sentiments expressed by the students on courses, facilities and tutors (Binali et al., 2009).

2.3 Types of Opinion

An opinion can be either a regular or comparative opinion. This section elaborates the types of opinion in detail.

2.3.1. Regular Opinion

A regular opinion is referred to as an opinion that can be categorized as explicit (direct) and implicit (indirect) opinion (Liu, 2006). If an opinion was expressed directly on an object or a feature, it is called a direct opinion. On the other hand, if an opinion was expressed indirectly on an object or a feature, it is called an indirect opinion. In the case of indirect opinions, opinions on objects are expressed based on their effects on some other objects (Liu, 2012). Consider the following two sentences:

Sentence 7: ‘*This car has good mileage*’

Sentence 8: ‘*After taking this medicine, my joint felt better*’

In sentence seven, the opinion holder expressed a direct opinion on a car. Sentence eight describes a desirable effect of the medicine on the joint, which indirectly presents a positive opinion about the medicine.

2.3.2. Comparative Opinion

A comparative opinion expresses a relation of similarities or differences between two or more objects and /or a preference of the opinion holder based on some shared features of the objects (Jindal & Liu, 2006a). For example, consider the following sentence that is expressing a comparative opinion on two digital cameras, namely, Canon G2 and Canon G3.

Sentence 9: '*Canon G3 is better than Canon G2*'

2.4 Different Levels of Semantic Analysis

This section discusses different granularity levels of opinion mining. In general, opinion mining has been investigated mainly at three granularity levels, namely, document-level, sentence-level and feature-level (Zhang & Liu, 2014).

2.4.1 Document-level (Review- level) Semantic Classification

Document-level opinion mining determines an overall opinion on an object (Liu, 2012). The objective of this analysis is to classify a whole opinion document either as positive or negative (Moghaddam, 2013).

Problem Definition: Given an opinion document *D* evaluating a target object *O*, determine the overall sentiment *s* of the opinion holder h_k about the object *O*, i.e., determine *s* expressed on object GENERAL in the quintuple

(O, GENERAL, s, h, t),

Where the object *O*, opinion holder *h*, and time of opinion *t* are assumed known or irrelevant (Liu, 2012).

Assumptions: Each opinion document focuses on a single object

Each opinion document contains opinion from a single opinion holder

For example, the classification of a review into positive or negative categories is called document-level semantic classification because it considers the whole review as the basic information unit. It assumes that an opinion document expresses opinions on a single object/product. Therefore, it is not suitable to compare or evaluate multiple products (Liu, 2012). Early opinion mining research focused on sentiment classification at the document-level, including Turney (2002) and Pang et al. (2002).

2.4.2 Sentence-level Semantic Classification

Semantic classification applied to individual sentences is called sentence-level semantic classification.

Problem Definition: Given a sentence *x*, determine whether *x* expresses a positive, negative, or neutral opinion (Liu, 2012).

Assumption: A sentence expresses a single sentiment from a single opinion holder (Liu, 2012)

Sentence-level semantic classification identifies whether each sentence expresses a positive, negative, or neutral opinion. However, each sentence of a review document cannot be assumed to be opinionated. Therefore, it is necessary to categorize a sentence as opinionated (subjective) or not opinionated (objective), which is called subjectivity classification (Zhang, 2012). Subjectivity classification determines whether a sentence is expressing factual or opinionated information (Wiebe, Bruce, & O'Hara, 1999). Objective sentences express factual information, while subjective sentences encode subjective views and opinions. The identified subjective sentences are then classified as positive or negative. As a result, it consists of two tasks:

Task 1: identify if a sentence is opinionated (subjective)

Task 2: determine the polarity of sentence (positive or negative)

Hatzivassiloglou and Wiebe (2000), Kim et al. (2004), and Riloff and Wiebe (2003) among others focused on sentence-level semantic classification. Both document-level or sentence-level analyses are useful in many cases, however, they are often inadequate for many applications as they are unable to provide detailed information in the form of opinion targets and their associated semantics (Liu, 2006).

2.4.3 Feature-level Semantic Classification

Although semantic classification at document-level and sentence-level is useful in many applications, however, it does not explore positive and negative comments on different features of a target object. A positive evaluation document does not mean that the opinion holder likes every feature of the object. Likewise, a negative opinion evaluative document does not mean that the opinion holder has negative opinions on every feature of the object (Zhang, 2012). For instance, the reviewers write both positive and negative features of products in the product domain, however, the overall sentiment on the product could be positive or negative. To overcome this shortcoming, more fine-grained opinion analysis is required to investigate opinion at feature-level. As a result, feature-based opinion mining was introduced by Hu and Liu (2004) to extract and summarize people's opinions expressed on objects or features of the objects.

Feature-level semantic classification performs fine-grained analysis by focusing on opinions themselves instead of language constructs, such as documents, sentences, clauses, or phrases. In this analysis, an opinion is considered to be a sentiment (positive or negative) and a target of the semantic. The objective of the analysis is to identify semantic on objects and/or their features.

2.5 Feature-based Opinion Mining

Mostly, opinion mining systems investigate opinions from a large collection of opinion holders. The opinion of a single person is not sufficient for decision-making, indicating the need for a detailed summary from numerous online reviews to support the decision-making process. Feature-based opinion mining addresses the needs for detailed summary (Moghaddam, 2013).

The objective of feature-based opinion mining is to (i) identify product features that have been commented on by opinion holders, (ii) determine whether the comments are positive or negative and (iii) finally produce a structured summary. The main goal of feature-based opinion mining is to produce a feature-based summary from multiple reviews (Liu, 2012). For example, consider the following sentence:

Sentence 10: "*I bought a canon powershot G3 camera yesterday, and its battery life is excellent, but it was a bit expensive*"

Feature-based opinion mining identifies that the opinion holder expresses a positive opinion on the feature '*battery life*' and a negative opinion on the feature '*price*' of the '*Canon PowerShot G3*' in sentence 10 above.

A model comprises of an object *O* and a set of opinions on the object *O* can be defined, which is called feature-based opinion mining model. In this model, an object *O* is described with a finite set of features, $F = \{F_1, F_2, \ldots, F_n\}$, which includes the object itself. Each feature $f_i \in F$ can be expressed with a finite set of opinion words or phrases, $W = \{W_1, W_2, \ldots, W_n\}$ for the total of 'n' features of the object *O*. In an evaluative document *D*, which evaluates the object *O*, an opinion holder *j* comments on a subset of the features $S_j \subseteq F$. For each feature $f_k \in S_j$, the opinion holder *j* comments on using a word or phrase from W_k to describe the feature, and then expresses a positive, negative or neutral opinion on f_k (Liu, 2006).

Similarly, a model of opinion document can be described based on the feature-based opinion mining model. In opinion document model, an opinionated document *D* contains opinions on a set of objects $\{O_1, O_2, \ldots, O_n\}$ from a set of opinion holders $\{h_1, h_2, \ldots, h_p\}$. The opinions on each object O_i are expressed on the object itself and a subset O_{id} of its features (Zhang, 2012).

There are three main steps in feature-based opinion mining for a given evaluation document *D*, which contain opinions on an object *O*. They are described below:

2.5.1. Step 1 - Identifying Object Features

The extraction of features of a target object *O* that have been commented on by an opinion holder *j* in an evaluation document *D* is the step one of the feature-based opinion mining. In the digital camera domain, picture quality, lens, battery and viewfinder are prominent features. For instance, *'battery life'*, *'flash'*, '*LCD display*' and '*self-timer*' are the features of Photosmart 435 digital camera as shown in Figure 2.1.

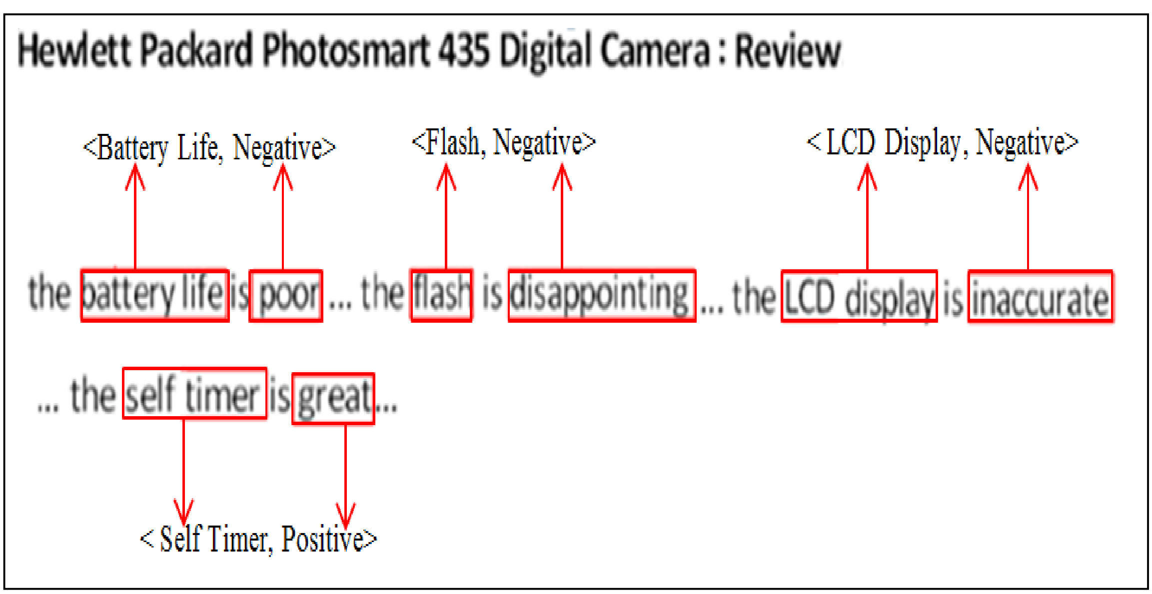


Figure 2.1: Identification of Features and Opinion Orientation

2.5.2. Step 2 - Determining Opinion Orientations

The second step is to identify whether the opinions on extracted features are positive, negative or neutral. In Figure 2.1, the opinion orientation on the features *'battery life'*, *'flash'*, *'LCD display'* is negative, whereas opinion orientation on feature '*self-timer*' is positive. The identification of features and opinion orientation on these features are presented in Figure 2.1.

2.5.3. Step 3 - Summarization and Visualization of Opinion Mining Results

The last step addresses the summarization and visualization of the overall opinion about a target object. Due to a large number of online reviews, some form of opinion summary is needed (Hu & Liu, 2004). There are many ways to present the summary, such as products or feature-based summary. Feature-based opinion mining uses features as the basis for an opinion summary.

Generally, the information discovered in step one and two are stored in database tables. Then, visualization tools are applied to see the opinion summary in different ways, e.g. bar chart or pie chart to gain insights of people's opinions (Zhang, 2012). In the early works of feature-based summarization, researchers focused on opinion summarization in the traditional fashion, i.e. displaying textual summary, which provides a quick overview of people liking and disliking a product or service (Liu, 2012). However, textual summary is not quantitative but only qualitative and is usually not suitable for analytical purposes (Liu, 2010). To overcome this weakness, the recent work of opinion summarization utilized visualization techniques to present users a concise, quantitative and visual view of an opinion summary (Zhang, 2012). Users can interact with visualization to obtain decision-oriented information conveniently.

The feature-based summary can be textual or non-textual. Figure 2.2 shows the opinion summary of customers' reviews on a particular digital camera, 'digital_camera_1', where '*CAMERA*' represents the camera itself (the root node in the object hierarchy). Customers' opinions on features '*picture quality*' and '*size*' (non-root nodes in the object hierarchy) of the camera are also presented in Figure 2.2. The digital_camera_1 was discussed positively in 125 reviews, whereas only seven reviews commented negatively. *'picture quality'* and *'size'* are important features of the camera. The *'picture quality'* of the 'digital_camera_1' received 123 positive opinions and only six negative opinions. The <individual review sentences> refers to the specific sentences that give the positive or negative comments about the feature.

*Digital_camera_*1:
CAMERA:
Positive: 125 <individual review sentences>
Negative: 7 <individual review sentences>
Feature: **picture quality**
Positive: 123 <individual review sentences>
Negative: 6 <individual review sentences>
Feature: **size**
Positive: 82 <individual review sentences>
Negative: 10 <individual review sentences>
…

Figure 2.2: Feature-based Textual Opinion Summary (Hu & Liu, 2004)

A non-textual opinion summary is shown in Figure 2.3 in which two digital cameras (digital camera 1 and digital camera 2) are compared based on customers' opinions. Bars with different colors encode cameras (the root node in the object hierarchy) and significant features of the cameras (non-root nodes in the object hierarchy). The height of a bar above and below the x-axis reflects the number of positive and negative comments on a particular feature, respectively. It can be concluded from Figure 2.3 that digital camera one received more positive comments than digital camera two on features: '*picture*', '*battery*', '*zoom*' and '*size*'.

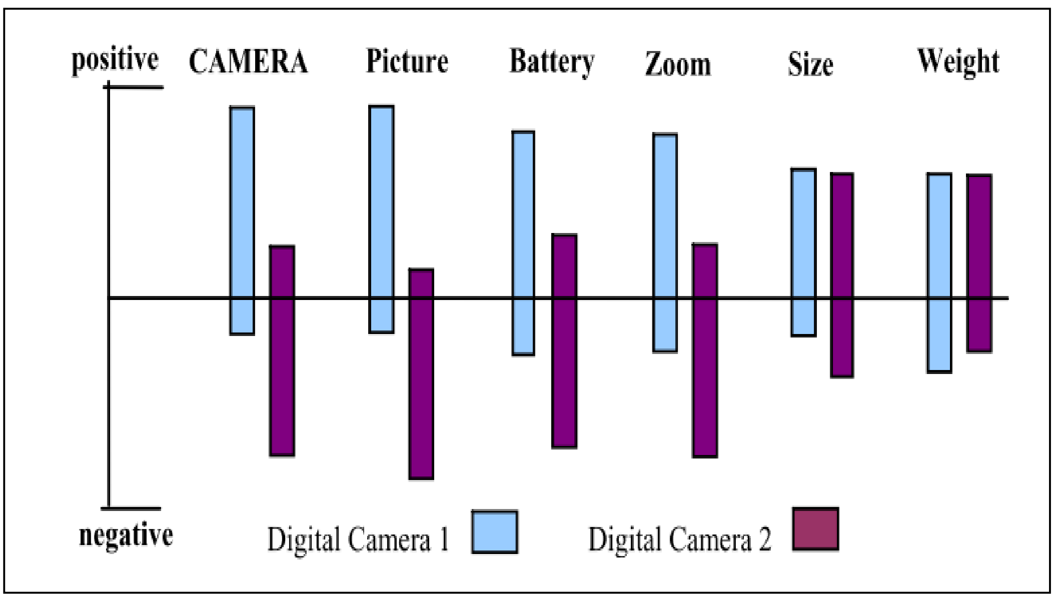


Figure 2.3: Feature-based Non-Textual Opinion Summary (Liu, Hu, & Cheng, 2005)

2.6 Opinion Mining Objective and Tasks

The objective of opinion mining is to discover all opinion quintuples (o_j, f_{jk}, s_{ijkl}, h_k, t_l) from a given opinion document *D* (Liu, 2006). The main tasks of opinion mining are derived from the components of the tuple (Liu & Zhang, 2012) are shown in Figure 2.4:

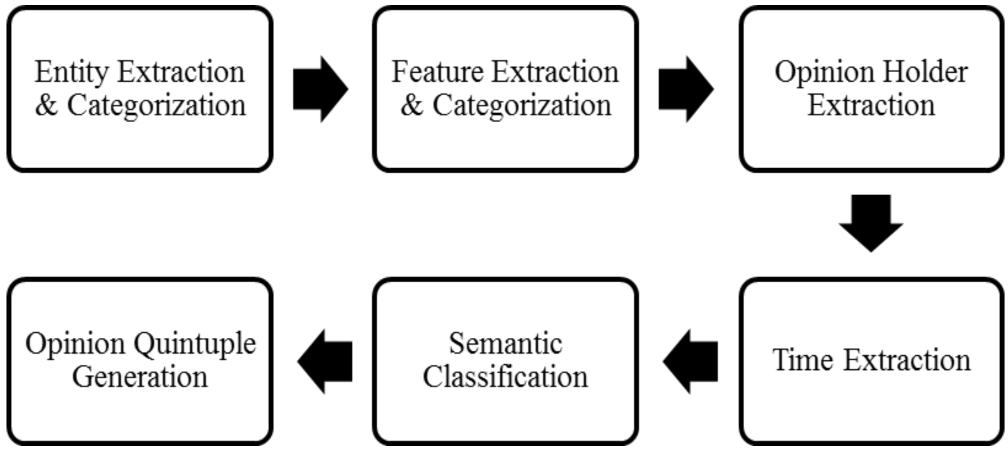


Figure 2.4: Opinion Mining Tasks

2.6.1 Object Extraction and Categorization

The target objects are extracted from a given opinion document *D* on which opinion holders express the semantic. Opinion holders often refer to the same object with different names, for instance, Motorola mobile may be expressed as Mot, Moto, and Motorola. Object categorization converts the object's synonyms into a single object.

2.6.2 Feature Extraction and Categorization

This task extracts products' explicit and implicit features. Opinion holders normally mention the same feature with different names, for instance, picture may be expressed as photo, image, and pics. Like object categorization, feature categorization groups together the same features expressed with their synonyms.

2.6.3 Opinion Holder Extraction

It extracts the opinion holder that expresses the opinion about a product or feature. The holder of an opinion is a person or organization that expresses a specific opinion on a product or feature.

2.6.4 Time Extraction

It extracts the time when the review was written by the opinion holder. Time is an important attribute of an opinion, since users are interested to know the customers' opinion trend movement with time.

2.6.5 Semantic Classification

Semantic classification classifies each opinion word as positive, negative or neutral.

2.6.6 Opinion Quintuple Generation

It produces all opinion quintuples (o_j, f_{jk}, s_{ijkl}, h_k, t_l) expressed in the opinion document *D* based on the results of the above tasks.

2.7 Phases of Opinion Mining

Opinion mining can be divided into three phases; namely, pre-processing, opinion mining, and post-processing as shown in Figure 2.5. These phases are discussed below:

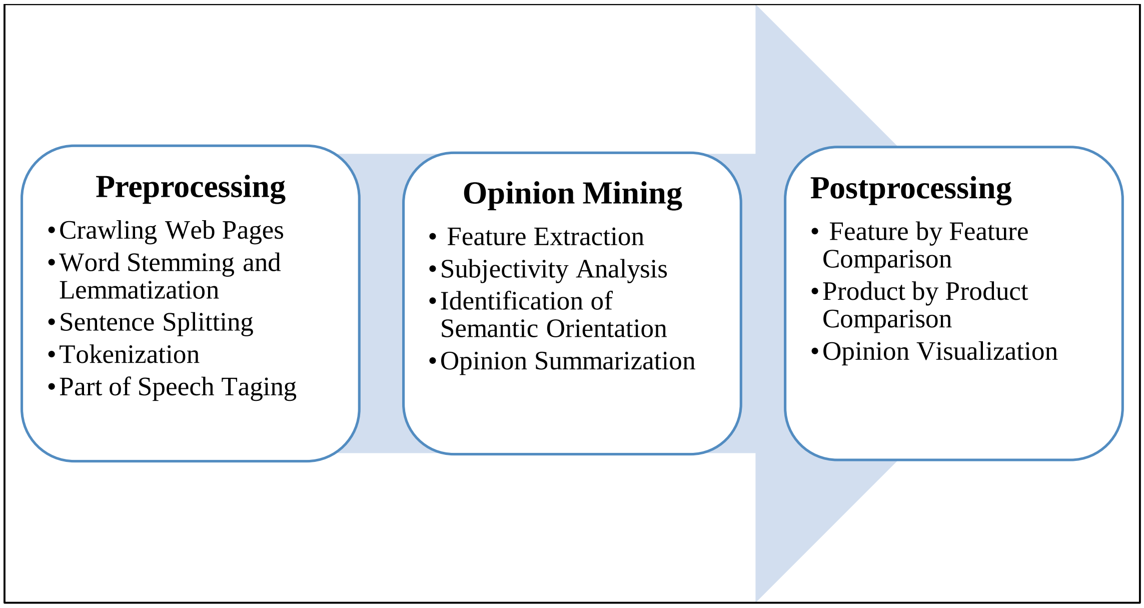


Figure 2.5: Opinion Mining Phases

2.7.1 Pre-Processing Phase

In this phase, different processes are applied to convert reviews in a structured format. The objective of pre-processing is to prepare text of the reviews for opinion mining. The most common processes in this phase are the extraction of online reviews, word stemming and lemmatization, sentence splitting, tokenization, and part of speech (POS) tagging. Common pre-processing steps are described below:

2.7.1.1 Extraction of Online Reviews

This process extracts online reviews from different online platforms. There are two approaches for the extraction of reviews; (i) wrapper induction and (ii) automatic pattern finding (Liu et al., 2005). In wrapper induction approach, users manually label a set of reviews and extraction rules are extracted from the labelled data. Then, these rules are used to extract reviews from other pages. The automatic pattern finding approach automatically identifies patterns from a page and then these patterns are employed to extract reviews from other pages. Both these approaches are for websites that display reviews according to some fixed layout templates (Liu et al., 2005).

2.7.1.2 Word Stemming and Lemmatization

Review documents contain inflectional forms (i.e., car, cars, car’s, cars’) and derivationally related forms (i.e., democracy, democratic, democratization) of a word for grammatical reasons. In information retrieval and opinion mining, stemming and lemmatization are used to reduce these forms to a common base form (Manning, Raghavan, & Schütze, 2008). For instance, in the following example, different forms of the word ‘*camera*’ are converted into its root form, that is, camera using word stemming and lemmatization.

$$\text{Camera, Cameras, Camera's, Cameras'} \Rightarrow \text{Camera}$$

However, there is a slight difference between word stemming and lemmatization. Stemming usually refers to a process that chops off the ends of a word, whereas lemmatization usually converts words to their roots with the use of a vocabulary and morphological analysis of words (Manning et al., 2008).

2.7.1.3 Sentence Splitting

Sentence splitting is also known as sentence segmentation. It is a process in which text of the reviews is fragmented into individual sentences. Sentence splitters are often integrated into tokenizers, but some separate tools are also available (Herold, Lemnitzer, & Berlin, 2012).

2.7.1.4 Tokenization

This process divides reviews' text into a sequence of simple tokens such as numbers, punctuation and words. It generates different tokens of a sentence.

2.7.1.5 Part of Speech (POS) Tagging

POS tagging is a process that takes tokenized text as input and associates a part of speech tag (POS tag) with each token. The POS tag of a word is a linguistic category that is defined by its syntactic or morphological behavior (Moghaddam, 2013). Common POS categories in English grammar are: noun, verb, adjective, adverb, pronoun, preposition, conjunction, and interjection. For instance, the POS tag of the sentence '*the pictures are great*' is shown below where DT presents a delimiter, NNS encodes a noun plural, VBP reflects a verb, and JJ represents an adjective:

DT /The /NNS pictures /VBP are /JJ great

After applying suitable pre-processes on the reviews, opinion mining is applied.

2.7.2 Opining Mining Phase

Opinion mining phase can be divided into three sub-phases, that is, feature extraction, sentiment orientation and opinion summarization (Ding et al., 2008), as shown in Figure 2.6. These sub-phases are briefly discussed in the following section.

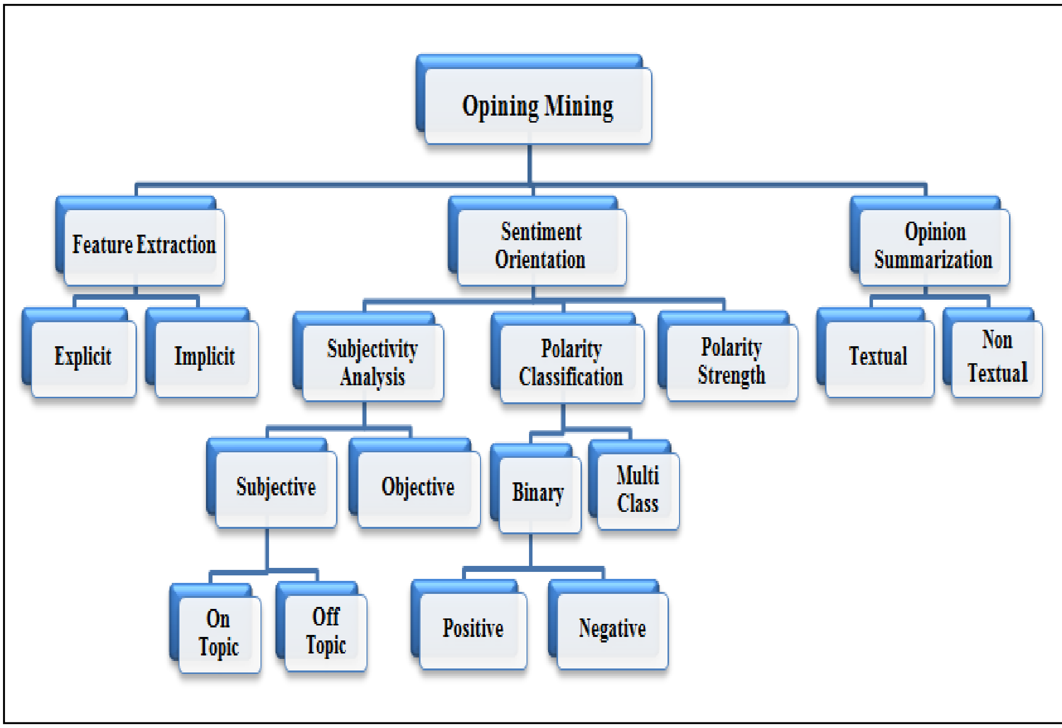


Figure 2.6: Sub-Phases of Opinion Mining

2.7.2.1 Feature Extraction

The identification of the product's features on which customers express an opinion is called product feature extraction.

2.7.2.2 Identification of Semantic Orientation (Polarity/ Classification)

Other names for sentiment orientation are opinion orientation, semantic classification or polarity. The identification of semantic orientation classifies each opinion word as positive, negative or neutral (i.e. opinion orientations). In practice, neutral is often interpreted as no opinion. It can be divided into three steps; (i) subjectivity analysis, (ii) semantic polarity classification, and (iii) polarity strength identification (Binali et al., 2009; Esuli et al., 2005) as discussed below:

A. Subjectivity Analysis

Subjectivity is the linguistic expression of somebody's opinion, sentiment and emotion (Ganesan & Kim, 2008). Subjectivity analysis determines whether a given document expresses an opinion or not (Moghaddam & Ester, 2013). It is the computational study of affect, opinions, and sentiments expressed in blog, review website, editorials, newspaper articles (Ganesan & Kim, 2008). The goal of subjectivity is to distinguish subjective sentences from objective sentences. Subjective sentences depict opinions, evaluations, sentiment, appraisal or emotions (Wiebe, Wilson, Bruce, Bell, & Martin, 2004), while objective sentences express some factual information about the world. Consider the following two sentences:

Sentence 11: '*Today, I bought Canon PowerShot G3*'

Sentence 12: '*The camera is good*'

Sentence 11 is expressing factual information, thus is an objective sentence, while sentence 12 is depicting someone's opinion, so is a subjective sentence. Subjective sentences can be categorized as on-topic and off-topic. If a subjective sentence describes a positive or negative semantic towards a specific topic then it is on-topic (relevant text) otherwise it is off-topic (irrelevant text) (Jijkoun, Rijke, & Weerkamp, 2010).

B. Semantic Classification

Semantic classification is also known as opinion classification, semantic orientation and sentiment polarity. It categories sentences/documents/features into positive, negative and neutral classes or on a numerical scale based on the semantics expressed by the opinion holders. Specifically, it classifies each subjective review sentence as positive or negative. The semantic classification of a word indicates the direction of deviation of the opinion word from the norm for its semantic group (Hatzivassiloglou & Wiebe, 2000). Words that encode a desirable state, such as beautiful, awesome, have a positive orientation, while words that represent undesirable states, such as disappointing, have a negative orientation (Ding et al., 2008). Consider the following review sentences.

Sentence 13: '*I bought a new camera yesterday. It was a bit expensive, but the battery life is very good.'*

In the above review sentence, the opinion word '*expensive*' represents a negative opinion on the feature '*price*', while the opinion word '*Good*' represents a positive opinion on the feature '*battery life*'.

There are two main types of semantic polarity classification: binary (bi-polar) and multi-class (fine-grained). In binary classification, the semantic of each sentence is classified into positive or negative classes, while in the case of multi-class; the classification is done on a scale (1 to n). If semantic takes categorical values (positive and negative), then it becomes a classification problem. If semantic takes numeric values or ordinal scale within a given range (1 to n), then it is a regression problem (Liu, 2012).

C. Polarity Strength Identification

The strength of an opinion word refers to how a text expresses a positive or negative opinion, i.e. weakly, mildly, or strongly (Lo & Potdar, 2009), for instance, the opinion word *'excellent'* is more positive than the opinion word *'good'*.

2.7.2.3 Opinion Summarization

Opinion summarization is orthogonal to feature extraction and semantic classification. Opinion summarization is the task of producing a sentiment summary from the information discovered from the previous processes. An opinion summary can be textual or non-textual (graphical) (Somprasertsri & Lalitrojwong, 2010).

2.7.3 Post-Processing Phase

In the post-processing phase, different graphical data visualization techniques are applied to present feature-by-feature, product-by-product comparisons and summarization of customers' opinions visually, which can enhance the user's ability to understand the customers' feedback effectively and efficiently. The high-level visualization of opinion mining systems can assist users to perform the comparison of different products/features at a single glance. This comparison is good for both customers as well as for manufacturers. A customer can consider the strengths and weaknesses of each product to make a decision about the purchase of a product. On the other hand, a manufacturer can discover the shortcomings of their products over different competitive products and improve them.

2.8 Performance Evaluation of Opinion Mining System

The performance of an opinion mining system can be evaluated by measures such as accuracy, precision, and recall if the ground truth (manually calculated values) is available. However, such ground truth is not available in real life datasets. In some cases, human judges manually create a set of true features list after reading a set of reviews. This set is called "gold standard" (Moghaddam, 2013), which is then used to evaluate the performance of opinion mining systems.

2.8.1 Accuracy

Accuracy expresses the degree of correctness by comparing an extracted value to an actual value (gold standard). It can be calculated with equation 1 given below:

$$Accuracy = \frac{Extracted\ Value}{Actual\ Value} * 100 \tag{1}$$

In this thesis, the metric accuracy is used to measure the effectiveness of the proposed review and feature ranking methods. The accuracy of these methods is described in Chapter 5.

2.9 Review Format

A review is a subjective text containing a sequence of words describing opinions of a reviewer regarding a specific item (Moghaddam, 2013). There are three types of reviews formats available on the Web (Liu, 2006) and are discussed below:

2.9.1 Pros and Cons Format

In this format, the reviewers express pros and cons of a product in the form of full sentences separately, for instance, Cnet.com practices this format as shown in Figure 2.7.

Figure 2.7: Pro and Cons Format (www.cnet.com)

2.9.2 Pros, Cons, and the detailed review

In this format, reviewers express pros and cons of a product with a detailed review, for instance, Epinions.com exercises this format as shown in Figure 2.8.

Figure 2.8: Pro, Cons and Detailed Review Format (www.epinions.com)

2.9.3 Free format

In this format, reviewers express their reviews freely without separation of pros and cons in full sentences, for instance, Amazon.com utilizes this format (Figure 2.9). Amazon.com consists of annotations that depicts how many people found the review helpful, for instance, '*15 of 20 people found the following review helpful*'. Each review on Amazon comes with both a star rating-the number of stars assigns to the product by the author of the review and helpfulness votes. This format is the focus of the thesis.

36 of 39 people found the following review helpful
Cool [Toy]., March 13, 2003
By **parrhesiastes "ti esti"** (Dublin, Ireland) - See all my reviews
This review is from: **Canon PowerShot G3 4MP Digital Camera w/ 4x Optical Zoom (Electronics)**
Yep. This is my first Digital Camera, and what a 'toy' it is!!! I am a software engineer and am very keen into technical details of everything i buy, I spend around 3 months before buying the digital camera; and i must say, G3 worth every single cent i spent on it.
Just a little overview, PowerShot G3 is the flagship of Canon's PowerShot series and its an SLR-like camera, its 4 megapixel and (alsmost) full manual control gives the pictures a touch of brilliance. Whether you are a novice or an expert, its ease of use and functionality goes together. + you can have different kind of lens if you want + flashes, etc. As its 4MP, you might need bigger storage to store high quality images and recording movies (you can record 3 minutes of video). I am using Kingston 512MB CF which works great and is very fast. A good choice could be CF type II, Microdrives that can store around 1 GB of images.
Some things that i didn't like (but hey.. nothing is perfect).
* Main dial is not backlit.
* Lens visible in optical viewfinder. (well thats only for old-school die-hard optical viewfinder fans) use LCD instead which is brilliant and you can twist around too.
Overall I'm happy with my toy. :)

Figure 2.9: Pros, Cons and Detailed Review Format (www.amazon.com)

The extraction of features and semantic are very challenging in format two and format three as they contain long and complete sentences with large amount of irrelevant information.

Chapter 3 : Literature Review

The first section of this chapter provides a comprehensive review on state-of-the-art review quality prediction and ranking methods. In the second section, existing approaches for feature-based opinion mining are discussed. Current feature ranking parameters are highlighted in section three. Finally, the last section presents a survey of existing opinion visualization techniques.

3.1 Review Quality Prediction

The increasing availability of online reviews presents a challenging recommendation opportunity – to recommend or rank reviews according to the helpfulness (O'Mahony & Smyth, 2009). Different studies have been conducted in the literature to predict review quality. This section introduces existing review quality prediction and review ranking apporaches.

The objective of review quality prediction is to determine the quality, helpfulness, usefulness, or utility of each review (Zhang & Varadarajan, 2006). It is desirable to rank reviews based on quality while showing reviews to the user, with the most helpful reviews first (Liu, 2012). Most of the review quality prediction approaches adopted textual features that are based on text statistics such as number of adjectives and nouns, length of the review, counts of specific POS tags, opinion words, and the average length of a sentence. However, recent works (Chen & Tseng, 2010; Ghose & Ipeirotis, 2011;

Moghaddam et al., 2012a; O'Mahony & Smyth, 2009) focused on social features that are related to the author of the review and are extracted from social context, i.e. PageRank of the reviewer, number of past reviews by reviewer, and past average rating for the author, in conjunction with textual features. The problem of review quality evaluation has been formulated as a classification or regression problem using the observed features (textual and/or social features) (Liu & Zhang, 2012). In this context, helpfulness votes given to each review in the reviewing websites served as the ground-truth for both training and testing purposes.

Kim et al. (2006) included metadata feature to predict the quality of a review according to helpfulness votes in addition to textual features by investigating the impact of five feature classes: (i) Structural, (ii) Lexical, (iii) Syntactic, (iv) Semantic and (v) Meta-data on the helpfulness of a review. Their experimental results revealed that structural features (Sentential and HTML) and syntactic features did not show any improvement in review quality prediction. Moreover, uni-gram features surpass the bi-gram features and semantic features performed as well as standalone features. The authors concluded that the best performing features are review length, uni-gram, and users' rating. Similarly, the influence of three feature classes, namely, (i) Lexical Similarity, (ii) Shallow Syntactic and (iii) Lexical Subjectivity Clues on the helpfulness of a review was explored by Zhang and Varadarajan (2006). The lexical similarity and lexical features with subjectivity clues played a very minor role while the shallow syntactic features resulted in the most predicting power in review quality prediction. Likewise, Liu et al. (2008) studied the effects of the reviewer's expertise, the writing style of the review based on POS tags, the timeliness of the review, review length, the polarity of the review and users' rating on the quality of a review. The authors found reviewer's expertise, the writing style of reviews and the timeliness of the review to be imperative in review quality prediction.

The helpfulness of a review depends not just on its content but also on users' social contexts (Mizil, Kossinets, Kleinberg, & Lee, 2009). The relationship between the social context of reviewers and the accuracy of a text-based review quality predictor was examined by Lu et al. (2010), who found that the social context reflects the quality of reviewers that in turn affects the quality of their reviews. Their experimental results showed that the use of social context information can help to improve the accuracy of review quality prediction. A classification approach was proposed by O'Mahony and Smyth (2009) to classify helpful and non-helpful hotel reviews based on both textual and social features, that is, Reputation, Content, Social and Sentiment. The authors found that reputation played a strong role, followed by sentiment in review quality prediction, however, the social and content had very limited influence on the review quality prediction (O'Mahony & Smyth, 2009). Table 3.1 shows the summary of the research work which is discussed above. The specific features and sub-features adopted by these studies are also presented in Table 3.1.

Table 3.1: Review Quality Prediction using Textual and Social Features

Study	Features	Sub-Features	Significant Features
Kim et al. (2006)	Structural	Length (LEN): The total number of tokens	Length Unigram Star Rating
		Sentential (SEN): the number of sentences the average sentence length the percentage of question sentences the number of exclamation marks.	
		HTML (HTM): Two features for the number of bold tags <b> and line breaks .	
	Lexical	Unigram (UGR): The tf-idf statistic of each word occurring in a review.	
		Bigram (BGR): The tf-idf statistic of each bi- gram	
	Syntactic	The percentage of tokens that are nouns, verbs, adjective, and adverbs	
	Semantic	Products Features	
		Positive and negative opinion word	
	Meta Data	Star Rating	
Zhang & Varadarajan (2006)	Lexical Similarity	TF IDF	Shallow Syntactic Features
	Shallow Syntactic	Noun Numbers Model Verbs Interjection Comparative and superlative adjectives Comparative and superlative adverbs Wh-determiners, wh-pronouns, possessive wh- pronouns, wh-adverbs: wh-words	

	Subjectivity Clues	Subjective adjectives Strong subjective nouns	
Liu et al. (2008)		Reviewer Expertise Writing Style Timeliness Review Length Polarity of the review Users' Rating	Reviewer Expertise Writing Style Timeliness
O'Mahony & Smyth (2009)	Reputation features	The mean and standard deviation of review helpfulness over all reviews authored by the user The percentage of reviews authored by the user which have received a minimum of T opinions	
	Content features	the number of terms in the review text the ratio of uppercase and lowercase charac- ters to other characters in the review text the ratio of uppercase to lowercase characters in the review text an integer which captures whether the user has completed one, both or none of the optional liked and dis- liked parts of the review the number of optional personal and purpose of visit details that are provided by the user the number of optional review-template questions that are answered in the review	
	Social features	the number of reviews authored by the user the mean and standard deviation of the number of reviews authored by all users; the number of reviews submitted for the hotel	

		the mean and standard deviation of the number of reviews submitted for all hotels.	
	Sentiment features	the score assigned by the user to the hotel the number of (optional) sub-scores assigned by the user the mean and standard deviation of the sub-scores as- signed by the user; the mean and standard deviation of the scores assigned by the user over all reviews authored by the user; the mean and standard deviation of the scores assigned by all users to the hotel.	

The effect of subjectivity, users' rating and readability on the helpfulness (the log of the ratio of helpful votes to total votes received for a review) of a review was analyzed by Ghose and Ipeirotis (2007) by demonstrating that review subjectivity has a significant impact on the perceived helpfulness of a review. The review consisting of both objective and subjective sentences with an extreme rating (one star or five stars) were found to be more helpful by users. In another study, Ghose and Ipeirotis (2011) combined semantic features with text subjectivity to explore the influence of five feature classes: (i) meta data, (ii) reviewer characteristics, (iii) reviewer history, (iv) review readability, and (v) subjectivity on helpfulness (the ratio of helpful votes to total votes received for a review) of a review. The variables used for each feature class is shown in Table 3.2. The review subjectivity, readability and reviewer history have shown statistically significant correlations with the helpfulness of reviews and the review with extreme rating was considered more helpful by users. Similarly, the impact of informativeness, readability and subjectiveness of reviews on review quality prediction was evaluated by Liu et al. (2007). The findings spotlighted that the informativeness (word-level, product-feature-level) and readability features can improve the performance of classification, however, the subjectiveness features had no contribution. Similarly, three components of review quality: credibility, informativeness, and readability were examined in (Mackiewics & Yeats, 2014) by focusing on 11 review characteristics (See Table 3.2). Reviewer's prior experience with a similar product (credibility), product research (credibility), a general recommendation (informativeness), headings (readability), met expectations (informativeness) and a specific recommendation about the product (informativeness) were found to be significant in review quality prediction.

Structural, syntactic, semantic and meta data features were explored to identify their impact on review quality and the results highlighted that high performance feature combination is rating, helpful percentage and review number which resulted in 69% accuracy (Zhang & Zhang, 2014). On the other hand, Pana and Zhang (2011) studied the effects of review characteristics, product type, and reviewer characteristics on perceived review helpfulness and identified significant impact of review valence and length on review helpfulness. Four types of social context: (i) author context, (ii) rater context, (iii) connection context, and (iv) preference context were utilized in (Tang, Gao, Hu, & Liu, 2013) to investigate their impact on review helpfulness and the results showed that author context provides best results. Ganun et al. (2013) evaluated the impact of star ratings, text-based ratings (feature, sentiment) and average rating for predicting restaurant rating and their results exhibited that text ratings resulting in better predicting accuracy as compared to the predictions using the star ratings. The effect of textual and social feature on perceived helpfulness of was examined by Moghaddam et al. (2012) and their experiment showed that the combination of these features provide more accuracy in prediction of review helpfulness. Recently, business, review and reviewer characteristics were explored by Bakhshi et al. (2015) and their model showed that the activity level of a user, reviewer average rating, reviewer review count and active days have a significant relationship with the review quality. Moreover, they highlighted that longer and objective reviews are the main identifiers of high quality reviews. A summary of the research work discussed above is presented in Table 3.2 along with specific features and sub-features utilized in these studies.

Table 3.2: Review Quality Prediction using Textual, Social and Subjectivity Features

Study	Features	Sub-Features	Significant Features
Ghose & Ipeirotis (2007)		Subjectivity Score	Subjectivity Users' Rating
		Standard deviation of the subjective scores for each review	
		Users' Rating	
		Review readability	
		the difference between the date of data collection and the release date of the product	
Ghose & Ipeirotis (2011)	Meta-Data	Users' Rating Helpfulness votes	Users' Rating Reviewer History Subjectivity Readability
	Reviewer Characteristics	Reviewer Rank Top-10 Reviewer Top-50 Reviewer Top-100 Reviewer Top-500 Reviewer Real Name Nickname Hobbies	

		Birthday Location Web Page Interests Snippet Any Disclosure	
	Reviewer History	Number of Past Reviews Reviewer History Macro Reviewer History Micro Past Helpful Votes Past Total Votes	
	Readability	Length (Chars) Length (Words) Length (Sentences) Spelling Errors Automated Readability Index (ARI) for the review Gunning–Fog index for the review Coleman–Liau index for the review Flesch Reading Ease score for the review Kincaid Grade Level for the review Simple Measure of Gobbledygook score for the review	
	Subjectivity	Subjectivity Score Standard deviation of the subjective scores for each review	

Liu et al. (2007)	Informativeness	Sentence Level Word Level Product feature level	
	Readability	The number of paragraphs in the review The average length of paragraphs in the review The number of paragraph separators in the review	Word Level Product feature level Readability Features
	Subjectiveness	The percentage of positive sentences in the review The percentage of negative sentences in the review The percentage of subjective sentences (regardless of positive or negative) in the review	
Mackiewics and Yeats (2014)	Credibility	Relevant Role	Experience with prior model Product research General recommendation Headings Met expectations Specific recommendation
		Experience with prior model	
		Experience with brand	
		Experience with similar product	
		Product Search	
		Testing	
	Informativeness	General recommendation	
		Specific recommendation	
		Product value to cost	
		Met expectations	
	Readability	Headings	
Zhang and Zhang (2014)	Structural Features	Sentence count Token count Token per sentences	Rating

	Syntactic Features	Noun percentage Verb percentage Adjective percentage Adverb percentage	Helpful percentage Review number
	Semantic features	Reviewer ranking Helpful percentage Review no	
	Meta data	Rating Review age	
Tang et al. (2013)	Author context		Author context
	Rater Context		
	Connection Context		
	Preference Context		
Ganun et al. (2013)	Text rating	Topics	Text rating
		Semantic	
	Star rating		
	Average rating		
Moghaddam et al. (2012)	Textual Features	number of tokens number of sentences ratio of positive and negative sentiment words ratio of verbs, ratio of adverbs	the combination of textual and social features

	Social Features	number of past reviews by the reviewer in-degree and out-degree of the reviewer PageRank score of the reviewer	
Bakhshi et al. (2015)	Business	Business Stars Business active days Business review count Median income Education bachelor +	review words reviewer average stars reviewer review count active days
	Review	Star Polarity Subjectivity Review words	
	reviewer	Reviewer average stars Reviewer review count Active days	

The quality of reviews was assessed based on the information quality framework that consists of nine information quality dimensions: (i) Believability, (ii) Objectivity, (iii) Reputation, (iv) Relevancy, (v) Timeliness, (iv) Completeness, (vii) Appropriate Amount of Information, (viii) Ease of Understanding, and (xi) Concise Representation (Chen & Tseng, 2010). Table 3.3 shows the features used in each quality dimension. The objectivity and appropriate amount of information dimensions achieved superior evaluation performances than others. Huang et al. (2009) looked at the review quality evaluation on the basis of the behavior of review authors by utilizing user reputation, seller degree for credibility, and expertise degree for assessing review quality. Their results highlighted a strong relationship between these features and review quality, hence suggesting that the users' information derived from the analysis of their transactions facilitates in assessing review quality.

Table 3.3: Review Quality Evaluation using Information Quality Framework (Chen & Tseng, 2010)

Information Quality Dimensions	Features	Significant Dimensions
Believability	The product rating deviation of a review.	Objectivity Appropriate Amount of Information Concise Representation
Objectivity	The number of opinion sentences, positive sentences, negative sentences, and neutral sentences in a review.The percentage of opinion sentences, positive sentences , negative sentences, and neutral sentences in all sentences of a review.The percentage of positive sentences and negative sentences in all opinion sentences of a review.The cosine similarity between the tf-idf vectors of a review and the product description	
Reputation	The number of reviews written by the reviewer The ranking of the reviewer.	
Relevancy	The number of times the product name, brand names, website names, and other product names are mentioned in a review.The percentage of the product name, brand names, website nam es, and other product names among all these name entities in a review.The number of opinion sentences containing the product name, brand names, website names, and other product names in a review. The percentage of opinion sentences containing the product name, brand names, website names, and other product names in all opinion sentences.	
Timeliness	The degree of duplication of a review	

	The interval (in terms of the number of days) between the current review and the first review of the product	
Completeness	The number of different product features, brand names, websites, and product names mentioned in a review.	
Appropriate Amount of Information	The number of product features, opinion-bearing words, words, sentences, and paragraphs in a review. The average frequency of product features in a review The number of sentences that mention product features in a review	
Ease of Understanding	The number of misspelled words in a review The average document frequency of review words. The position of the first opinion sentence in a review The moving-average type/token ratios in a review	
Concise Representation	The average length of sentences and paragraphs in a review. The average number of sentences and opinion sentences in each paragraph of a review.	

Moghaddam and Jamali (2012) argued that the quality of a review may not be the same for different users and to address this issue they proposed a personalized review quality prediction model for the recommendation of helpful reviews. The proposed latent factor model showed superior performance than the state-of-the-art approaches using textual and social features, which confirmed that the helpfulness of a review is not the same for all users. Existing approaches also do not consider that the top few high-quality reviews may be highly redundant and repeating the same information (Tsaparas et al., 2011). To address this issue, a model for the selection of a comprehensive set of high-quality reviews that cover many different features of a target object and also different viewpoints of the reviews was proposed by Tsaparas et al. (2011).

Studies were also conducted to determine the relationship between different metadata features of reviews, e.g. review length, users' rating and helpfulness votes. Mudambi & Schuff (2010) analyzed the effects of rating, review length and product type on the perceived helpfulness of a review, and found that product type moderates the effect of users' rating and review length, which positively affects the helpfulness of the review. Similarly, the relationship between users' rating, review length and review readability was explored in Korfiatis et al. (2012). The helpfulness of a review was found to be affected by rating and readability. Additionally, the length of the review also has a positive relationship with users' rating.

In the literature, researchers ranked reviews based on different parameters such as number of features, number of opinion words and helpfulness votes. Tsur and Rappoport (2009) ranked dominant terms according to the respective frequency in reviews collection and the British National Corpus followed by ranking the reviews based on the number of dominant key terms. The empirical results exhibited that the proposed method outperformed term frequency-based method for book reviews. Similarly, Scaffidi et al. (2007) ranked reviews according to the number of features discussed in the review and their proposed method outperformed Feature-based Summarization System (FBS) developed by Hu and Liu (2004) in terms of precision. Reviews were ranked according to the number of features and opinion words in Eirinaki et al. (2012) and the results highlighted that the proposed method outperformed the conventional TF (term frequency) and TF-IDF (Term frequency-inverse document frequency) based methods. The lucene rank with temporal opinion quality was utilized in Miao, Li, and Dai (2009) to find relevant search results against users' queries. This union showed significant improvement over the conventional lucene rank method.

Reviews were classified into four categories: '*best review*', '*good review*', '*fair review*' and '*bad review*' based on whether reviews discuss many features of a target product by Liu et al. (2007). They authors considered the '*bad review*' category as a low quality class and all the other three categories were considered as a high quality class. Similarly, Chen and Tseng (2010) categorized reviews in five quality classes: '*high-quality*','*medium-quality*','*low-quality*','*duplicate*' and '*spam*,' according to the specification of review quality proposed by Liu et al. (2007), and the definition of spam reviews suggested by Jindal and Liu (2008). According to the authors, a high-quality review reflects complete, timely and sufficient opinion information about a target product to assist customers in making purchase decisions. A medium-quality review is relevant, however, it is not informative enough to influence customers' purchase decisions. The content of a low-quality review comprises little information about a target product, or the information is too objective to judge the value of the product. A review is considered as a duplicate if its content is similar to a review posted previously. Finally, a spam review only provides comments about product-irrelevant matters, or it may be an advertisement or a question-answer type of review.

It is to note that the review ranking approaches discussed above (Eirinaki et al., 2012; Miao et al., 2009; Scaffidi et al., 2007; Tsur & Rappoport, 2009) did not focus on users' preferences that define the important parameters according to the users' perspectives. For some users, helpfulness votes are more important than users' rating and for some users, the case is totally opposite. Therefore, users' preferences must be considered in order to fulfil customers' needs.

3.2 State-of-the-art Feature-based Opinion Mining Approaches

In a sentence, an opinion always has a target, which is often a feature or object. Therefore, it is imperative to identify opinion and its target (feature) from a sentence. The task of identifying the target of an opinion expression is called feature extraction, and is considered as an information extraction problem. This section discusses state-of the-art feature-based opinion mining approaches. There are three types of approaches for feature extraction, namely, frequency-based approach, relationship-based approach, and model-based approach as discussed below:

3.2.1 Frequency-based Approach

The earlier work on feature extraction utilized frequency-based approach in which high frequency nouns and noun phrases are considered as candidate features (Hu & Liu, 2004). This approach employed a set of manually or automatically extracted POS patterns to extract features (Moghaddam, 2013). Although, the approach is very simple and effective, however, it neglects many useful low frequency features and produces many non-features. Additionally, it is hard to port the approach for other datasets as it requires manual tuning of various parameters (thresholds).

The pioneer work on frequency-based approach was done by Hu and Liu (2004) who developed a system called FBS (feature-based summarization) by applying POS tags to identify frequent nouns and noun phrases as candidate features. The candidate features which do not appear in a specific order and are subsets of others were removed using two pruning methods resulting in 92% of precision. Popescu et al. (2005) introduced OPINE, an opinion mining system for the extraction of features and associated opinions. They tried to remove those noun phrases that may not be considered as features and its feature assessor component evaluates noun phrases by deploying Point-wise Mutual Information (PMI) (Turney, 2002) score between the phrase and its associated discriminators. A discriminator is an extraction pattern, which was used to find components or parts of a target object, for instance, the discriminator of camera object are '*great X*', '*has X*', '*comes with X*' , where X is a feature of the product. OPINE defined extraction rules based on the syntactic dependencies to find opinion words, and then its relaxation labelling technique classifies the orientation of opinion words. The authors reported 79% precision for the opinion phrase extraction of OPINE.

The frequency-based approach was improved by considering noun phrases that appeared in sentiment-bearing sentences or are in some syntactic patterns (Blair-Goldensohn et al., 2008). The authors also applied several filters to drop candidate features that are not associated with sufficient opinion words. Their findings showed 83% and 84% precisions for positive and negative comments, respectively. Ku, Liang and Chen (2006) considered the conventional TF-IDF (term frequency-inverse document frequency) scheme at document and paragraph levels with achievement of 77% and 74% precisions for opinion extraction tasks. Zhu, Wang, Tsou, and Zhu (2009) proposed a technique to extract multi-word features based on the frequency of multi-word term *t*, the length of *t*, and also other terms that contain *t* (Zhu et al., 2009). A set of candidate features was refined using a bootstrapping technique with a set of given seed features based on the idea of candidate's co-occurrence with the seeds. The proposed technique resulted in 82% of precision. Eirinaki et al. (2012) proposed a different approach for feature extraction by selecting the extracted nouns, which are associated with opinion words above a threshold. The proposed method showed superior performance than the conventional TF and TF-IDF based methods.

3.2.2 Relationship-based Approach

Relationship-based approach overcomes the weaknesses of the frequency-based approach by exploiting natural language relationships between opinions and their targets (features) based on the intuition that each sentiment expresses an opinion on a feature, and sentiments are often known or easy-to-find (Liu, 2012). These approaches extract low frequency features; however they produce many non-feature matches.

Hu and Liu (2004) also deployed this approach for the extraction of infrequent nouns. An opinion mining system, Opinion Observer was developed for the visual comparison of competitive products based on supervised pattern discovery algorithm in Liu et al. (2005). Association rules were extracted from POS tags in which actual features were replaced by a specific tag to extract both explicit and implicit features. Then to filter out non-predictive patterns some conditions were applied, resulting in 92% precision for a digital camera reviews. A dependency parser was utilized to identify dependency relations for feature extraction in Zhuang et al. (2006) achieving 66% precision for the movie review domain. Another similar approach was also deployed by Kobayashi, Iida, Inui and Matsumoto (2004) with a precision of 81% for the opinion extraction phase.

Pre-defined patterns and General Inquirer (GI) were utilized by Baccianella, Esuli and Sebastiani (2009) to extract nouns and noun phrases as candidate features and to predict the orientation of opinion phrases, respectively. The experimental results showed that the proposed methods outperformed the conventional BOW (bag of word) method. The idea of dependency parser was further enhanced by utilizing phrase dependency parser instead of a normal dependency parser for the extraction of noun and verb phrases as candidate features (Wu, Zhang, Huang, & Wu, 2009). In this work, a filter was also applied to remove unlikely features, resulting in 72% reduction of irrelevant candidate features. Existing exact matching methods based on syntactic structures of the features are unable to handle similar syntactic structure and therefore cannot be generalized for unseen data (Moghaddam, 2013). To overcome this problem, a tree kernel-based approach was proposed for the implicit exploration of syntactic sub-structure, sentiment polarity information and to calculate the similarity between two sub-structures (Jiang, Zhang, Fu, Niu, & Yang, 2010). The authors reported promising results for both sentiment expression extraction and sentiment classification. Du and Tan (2009) built a graph of features (noun phrases) and sentiments based on their co-occurrence in reviews followed by a graph clustering algorithm to find features highly related to sentiments, that outperformed the template extraction based approaches. Hai, Chang and Kim (2011) utilized a co-occurrence association rule mining approach to extract implicit features that resulted in considerable improvements over the PMI approach discussed by Turney and Littman (2003). Similarly, a double-propagation method based on the previous idea was proposed by Qiu, Liu, Bu and Chen (2011) for the extraction of both sentiment words and features simultaneously, utilizing syntactic relations that link opinion words and their targets by demonstrating 16% improvement in precision over the FBS.

3.2.3 Model-based Approaches

Model-based approaches were explored to overcome the shortcoming of frequency-based and relationship-based approaches, that is, the manual tuning of various parameters which makes them hard to port to another dataset or domain. The model-based approach addresses this weakness by automatically learning the model's parameters from the training data. There are two categories of the approach: supervised learning and unsupervised topic modelling techniques.

3.2.3.1 Supervised Learning Approach

Supervised learning information extraction methods are successfully applied to feature extraction, as feature extraction is considered as a special case of information extraction (Liu, 2012) that requires manually labelled data for training a model. The supervised learning approach overcomes the weaknesses of previous approaches; however, it still requires manually labelled data for training the models. The most prominent information extraction approaches are based on sequential learning such as Hidden Markov Model (HMM) and Conditional Random Field (CRF).

A lexicalized HMM model was applied to learn extraction patterns for feature and opinion expressions and the results demonstrated 77% precision for the feature-opinion pair orientation (Jin & Ho, 2009). Wong et al. (2011) integrated a probabilistic graphical model with HMM for extraction and grouping of features from several websites and layouts by predicting the label of each token as attribute-name, attribute-value or attribute-irrelevant. The authors reported 75% precision for digital camera reviews. Supervised learning approach was combined with pattern discovery to identify comparative sentences from news articles, consumer reviews of products and forum posts by Jindal and Liu (2006) resulting in 79% of precision. In another study, Jin et al. (2009) developed a system, called OpinionMiner in which POS tags were integrated with the HMM framework for feature and opinion extraction that showed significant performance improvement over the FBS in terms of feature extraction and opinion polarity classification. Similarly, HMM with topic modelling was utilized by Sauper, Haghighi and Barzilay (2011), where the HMM models the sequence of words with types (i.e. feature word and sentiment word) and their method outperformed the conventional TD-IDF method considerably.

Jakob and Gurevych (2010a) trained CRF on review sentences collected from different domains for a domain independent feature extraction and the approach improved the performance of Zhuang et al. (2006). Similarly, in Jakob & Gurevych (2010b), a CRF-based model was proposed with an achievement of performance improvement over Zhuang et al. (2006). Various experiments were conducted to compare the performance of different CRFs such as linear, tree, skip for the extraction of feature and opinions by Li, Han, Huang and Zhu (2010) with linear CRF model achieving the best performance for movie reviews. Similarly, Choi and Cardie (2010) also deployed CRF based on a sequential pattern mining technique and their experimental results achieved up to 92% precision for the opinion extraction. CRFs were employed for defining and filtering features in Shariaty (2011) resulting in better accuracy.

3.2.3.2 Topic modelling Approach

Researchers recently focused on topic modelling methods, which are an un-supervised learning technique based on the intuition that topics from topic models can represent features. It assumes that each document consists of a mixture of topics and each topic is a probability distribution over words (Liu, 2012). There are two basic models: PLSA (Probabilistic Latent Semantic Analysis) and LDA (Latent Dirichlet Allocation). A joint PLSA model ‘aspect-sentiment mixture model’, which is based on an aspect model and two sentiment models for positive and negative sentiments, respectively, was proposed in Mei, Ling, Wondra, Su and Zhai, (2007). The results demonstrated that the proposed model is effective for topic-sentiment analysis because it generates more useful topic-sentiment summaries for blog search than the blog opinion search engine (Opinmind). Titov and McDonald (2008a) proposed an extended topic model-based on LDA and PLSA for feature extraction from hotel reviews that focused on two types of topics: global and local. The global topics represent global property of the product in the review, whereas local topics represent the product features. The proposed model extracts features as well as clusters them into coherent topics which exhibited a significant improvement over standard topic models. The previous work was enhanced by deploying a set of maximum entropy classifiers, one per known feature to predict feature rating with 94% precision for one feature (i.e. location) (Titov & Mcdonald, 2008b).

Structured and unstructured PLSA with k-means clustering was utilized for feature extraction and clustering short comments by Lu, Zhai and Sundaresan (2009), and their results revealed that the accuracy of structured PLSA is better than unstructured PLSA. An un-supervised method based on latent semantic association (LaSA) for grouping candidate features was discussed in Guo, Zhu, Guo, Zhang, & Su (2009). First, they extracted all noun phrases as candidate features. Second, two alternative LaSA models were utilized to group the extracted features, where the first LaSA model deployed the page-independent context information to group features, and the second model employed page dependent layout information of text fragments. They achieved 88% precision in the digital camera domain. Zhao, Jiang, Yan, & Li (2010) proposed a LDA-based model which combined discriminative maximum entropy (Max-Ent) component with the standard generative component for identifying features and sentiments simultaneously with an achievement of up to 83% precision.

3.3 Feature Ranking

Overall product rating (star rating) is an important measure, however, different product features are important to different customers based on their usage patterns and requirements (Kunpeng et al., 2010). For instance, a digital camera that is overall ranked high may have poor battery life. Therefore, more detailed feature-based ranking is required to identify the features most liked and disliked by customers. Feature ranking is very useful in practice, as a large number of features may be extracted from a large dataset; however, users are often only interested in those important ones, which should be ranked high. It is also desirable to rank frequent features at top because they are more important than the infrequent ones (Lei et al., 2010).

There are several ways to assess the importance of each feature by assigning a weight. The most popular ones are: feature frequency (FF) and4 TF-IDF. Feature frequency is the common way to rank features. The idea behind this ranking is that if a feature is frequently mentioned in a dataset, thus it is an important feature and it should be ranked high. TF-IDF weighting is another scheme intended to attenuate the effect of terms that occur too often in a dataset. The TF-IDF consists of two parts; TF and IDF where TF represents the frequency of a term (feature) in a document and IDF reflects the number of documents in a dataset which contain the term.

$$tf - idf_{t,d} = tf_{t,d} \times idf_t \tag{2}$$

$$idf_t = Log \frac{N}{df_t} \tag{3}$$

Equation two shows the formula to assign a TF-IDF weight to a term *t* in document *d*. The formula to calculate the IDF value is presented in equation tree where *N* indicates the total number of documents and df_t represents the number of documents that contain term *t*. Existing feature ranking methods utilized users' rating, semantic polarity, feature frequency and opinion words. This section introduces existing work on feature ranking.

A search engine called Red Opal was introduced in Scaffidi et al. (2007) for product ranking based on features scores. The Red Opal identifies products' features and scores each product on each feature. Product rating and feature frequency were utilized in the system to compute product scores based on features mentioned in the reviews. The advantageous factor of the system is the estimation of a distinct score for each product on each feature. For instance, if there are 10 features for a certain product category (e.g. fiction books or ghost story), then 10 corresponding scores for each product in the category were estimated. The system surpassed the precision and efficiency of the FBS system proposed by Hu and Liu (2004). Similarly, the ranking of products based on features was discussed in Kunpeng et al. (2010). Features for each product category were extracted followed by the identification of four different types of sentences: positive subjective, negative subjective, positive comparative and negative comparative in reviews for product's ranking based on features. A weighted and directed feature-based product graph was then built that captures the sentiments expressed by customers in reviews in which nodes present products and edges describe comparative relation between products. A node weight reflects the ratio of the number of positive and negative subjective sentences whereas an edge weight corresponds to the ratio of the number of positive and negative comparative sentences. Finally, the importance of each product according to a particular feature was calculated using the RageRank algorithm that utilized nodes and edges weights. Experimental results showed that the proposed method achieved significant agreement with experts' evaluations.

Lei et al. (2010) ranked products' feature by feature importance, which is determined by two factors: feature frequency and feature relevance. Feature frequency is the occurrence frequency of a feature in a dataset, whereas feature relevance defines how likely a candidate feature is a correct feature. The authors applied Hyperlink-induced topic search (HITS) link analysis algorithm to compute feature relevance that resulted in better recall than Qiu et al. (2009). Similarly, an approach to rank product features according to their importance was discussed in Li et a. (2011). Features were extracted by exploiting the relationship between opinion words and product features, that is, the opinion words were adapted to extract product features followed by feature ranking according to the number of associated opinion words and product rating. Their results demonstrated the superiority of the proposed method over Qiu et al. (2009).

Eirinaki et al. (2012) introduced an opinion search engine AskUs by ranking features according to associated opinion words. A score called opinion score is assigned to each feature, which is increased by one for each associated opinion words. Features are ranked on the basis of their opinion scores such that the higher ranked feature is described by more opinion words. The opinion search engine resulted in better precision over the conventional TF and TF-IDF based methods.

An un-supervised opinion mining approach Opinion Digger, which extracts important products' features and determines the overall customers' satisfaction for each feature by estimating a rating in the range of 1 to 5 was proposed in Moghaddam & Ester (2010). In this work, features were ranked according to opinion words and the rating guideline provided by the review website (Epinions.com). The opinion words' rating associated with each feature was aggregated to estimate the feature rating that resulted in better accuracy than the FBS system (Hu & Liu, 2004).

Yang et al. (2010) highlighted that the existing feature ranking methods that utilized product rating to evaluate individual product features might be incorrect because the product rating provides evaluation for the entire product and is not an evaluation of a specific feature. To address this issue, the sentiment polarities of the opinion words, feature frequency and product rating were utilized to determine feature rank with results showing significant improvement over Scaffidi et al. (2007).

An opinion mining system was proposed by Ahmad & Doja (2012), that extracts product features and determines the strength of the opinions. A TF-IDF value for each extracted noun phrase was calculated and the noun phrases having TF-IDF values above a threshold were considered as relevant features. The system calculates two numeric scores of all features; one for positive evaluation and second for negative evaluation using Senti-

WordNet and the scores are summed up to calculate an overall score with an achievement of 92% precision.

3.4 Opinion Visualization Techniques

Opinion visualization techniques are required to communicate opinion mining results and facilitate the analytical reasoning process effectively. These techniques empower users to draw meaningful conclusions by providing a purposeful representation of the data and an appropriate starting point for the interactive exploration of attractive opinion patterns from a large collection of reviews (Wanner et al., 2009). The careful design of opinion visualization techniques is required to present customer's opinions with sufficient visual cues and different levels of abstraction (summarization), as this information has a significant impact on building a successful business (Wu et al., 2010). Different kinds of visualization techniques are suggested to visualize outcome of opinion mining, enabling further opinion analysis such as customers' behavior analysis, identification of customers' satisfaction, trust and loyalty over time, analyzing the relationships between demographic characteristics (e.g., age and gender), identifying the groups of customers with a similar opinion and correlations between the different features of the data set (Oelke et al., 2009; Wu et al., 2010). Every technique has its own level of abstraction, advantages and disadvantages. There are five types of opinion visualizations, namely, radial, hierarchical, graph, bar chart, and maps (Table 3.4). This section presents a comprehensive review of existing opinion visualization techniques.

Table 3.4: Existing Opinion Visualization Techniques

No.	Authors	Type of Visualization	Visualization
1.	Wu et al. (2010)	Radial	Opinion Wheel
2.	Gregory et al. (2006)	Radial	Rose Plot Variation
3.	Gamon et al. (2005)	Hierarchical	Tree Map
4.	Oelke et al. (2009)	Hierarchical	Visual Summary
5.	Chen et al. (2006)	Graph	Coordinated Graph
6.	Bjørkelund et al. (2012)	Graph	Coordinated Graph
7.	Filho et al. (2012)	Graph	Line Graph and Pie Chart
8.	Miao et al. (2009)	Graph	Line Graph and Pie Chart
9.	Guo et al. (2010)	Graph	Pie Chart
10.	Wang and Araki (2007)	Graph	Pie Chart
11.	Gamon et al. (2008)	Bar Chart	Glowing Bar
12.	Wanner et al. (2009)	Bar Chart	Bar Chart with Symbols
13.	Liu et al. (2005)	Bar Chart	Bar Chart
14.	Wang and Araki (2007)	Bar Chart	Bar Chart
15.	Dey and Haque (2008)	Bar Chart	Stacked Bar Chart
16.	Kongthon et al. (2011)	Bar Chart	Stacked Bar Chart
17.	Guo et al. (2010)	Bar Chart	Stacked Bar Chart
18.	Morinaga et al. (2002)	Map	Positioning Map

19.	Xu et al. (2011)	Map	Comparative Relation Map
20.	Bjørkelund et al. (2012)	Map	Google Map Variation
21.	Rohrdantz et al. (2012b)	Map	Pixel Map Calendar and Time Density Plot
22.	Hao et al. (2011)	Map	Pixel Cell-Based Sentiment Calendar and Geo Map
23.	Hao et al. (2013)	Map	Key Term Geo Map

3.4.1. Radial Visualization

Data in radial visualizations are arranged in a circular or elliptical fashion and is an increasingly popular visualization metaphor in visualization research (Draper, Livnat, & Riesenfeld, 2009). Opinion wheel (Wu et al., 2010) and rose plot variation (Gregory et al., 2006) employed radial visualization in opinion mining. Opinion wheel assimilates scatter plot inside an opinion triangle which is bounded by an opinion ring as shown in Figure 3.1. Here the opinion wheel provides users with an integrated view of multiple dimensions of opinion data, such as demographics and spatial information for online hotel customers' reviews. The opinion triangle presents positive, negative, and uncertainty opinion values on its negative (N), positive (P), and uncertainty (U) vertices. Customers' opinions and their semantic orientations were represented by a point and its position inside the triangle, respectively. The opinion wheel encodes the categories of different data dimensions with colored histograms on the opinion ring. The number of customers for each trip type was represented by the size and color of the histogram (Figure 3.1).

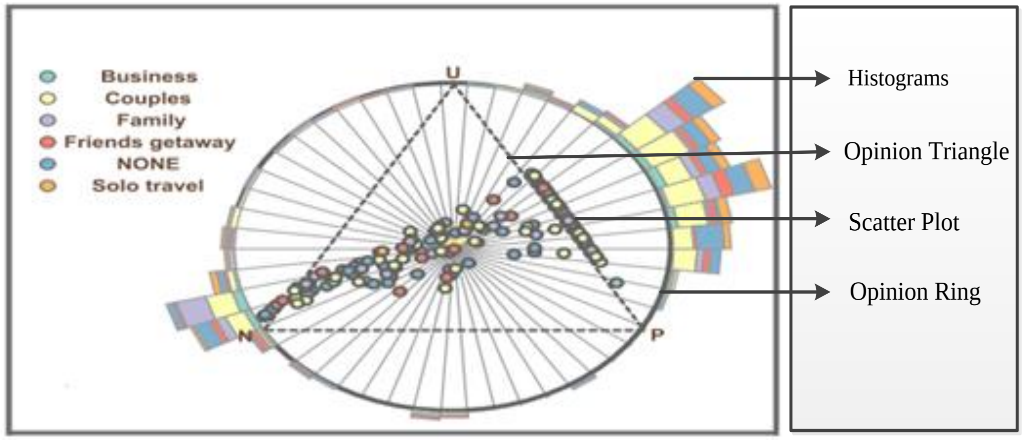


Figure 3.1: Opinion Wheel showing Customers' Opinions according to Trip Type (Wu et al., 2010)

The strengths of the opinion wheel are the use of familiar visual metaphors (positioning map, histograms) to visualize the results from complex opinion analysis, low level of abstraction and multidimensional view of data in an effective way; however, it is difficult to scale up the opinion wheel because of the limited space inside the opinion triangle.

A visual analytical tool was developed by employing a radial visualization to explore the sentiment contents in a large collection of documents (Gregory et al., 2006). The authors modified the rose plot originally used by Florence Nightingale (Nightingale, 1858) as shown in Figure 3.2. The first modification was the use of colors with different shades (light and dark) to represent emotion categories, whereas the second modification introduced the unit circle in the rose plot to display the mean (dotted line in Figure 3.2) and deviation values of opinions by drawing the appropriate rose plots outside (larger than mean) or inside (smaller than mean).

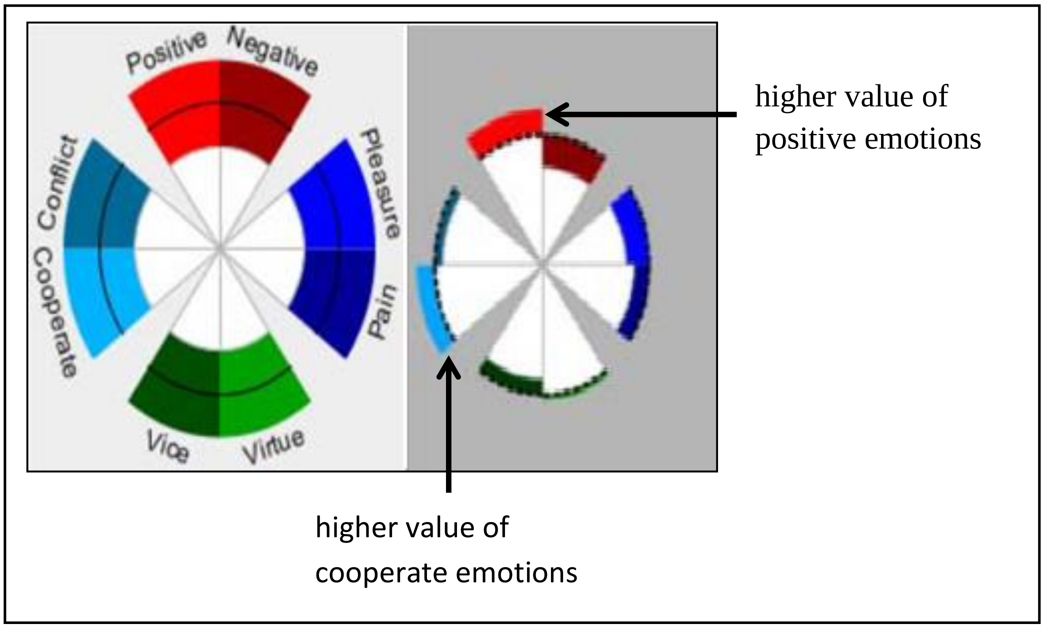


Figure 3.2: Rose Plot shows Emotion Categories and Deviated Values of Positive and Cooperate Categories (Gregory et al., 2006)

In this work, each document was assigned a score according to the eight emotion categories consisting of four concept pairs (positive-negative, pleasure-pain, cooperate-conflict, and virtue-vice). The glyph on the right hand side shows the higher value of positive and cooperate emotions than the mean value. The rose plot is a powerful visualization technique to analyse and compare a large collection of documents with respect to emotional categories. The authors also used histograms to present the number of documents in each group. The rose plot is aesthetically appealing, compact and easy to interpret visualization technique with enhanced comparison ability incorporating larger scores which draw more attention; hence, making it easier to compare semantic across large document groups.

3.4.2. Hierarchical Visualization

The hierarchical visualization was deployed in tree map (Gamon, Aue, Corston-Oliver, & Ringger, 2005) and visual summary report (Oelke et al., 2009). Gamon et al. (2005) showed the applicability of tree map to present customers' opinions on cars by developing Pulse, a prototype system. Sentences from reviews were aggregated into different clusters. Then, these clusters were rendered as boxes in the tree map. Each box was labeled with a keyword. The number of sentences in each cluster and average semantic on each cluster were encoded in the size and color of the boxes as shown in Figure 3.3. The color of boxes which ranges from green to red corresponds to the sentimental tendency, green for positive and red for negative. The visualization provides a high-level of abstraction of the overall sentiment associated with a target car, the most common topics (features), the most positive topics and the most negative topics at a glance.

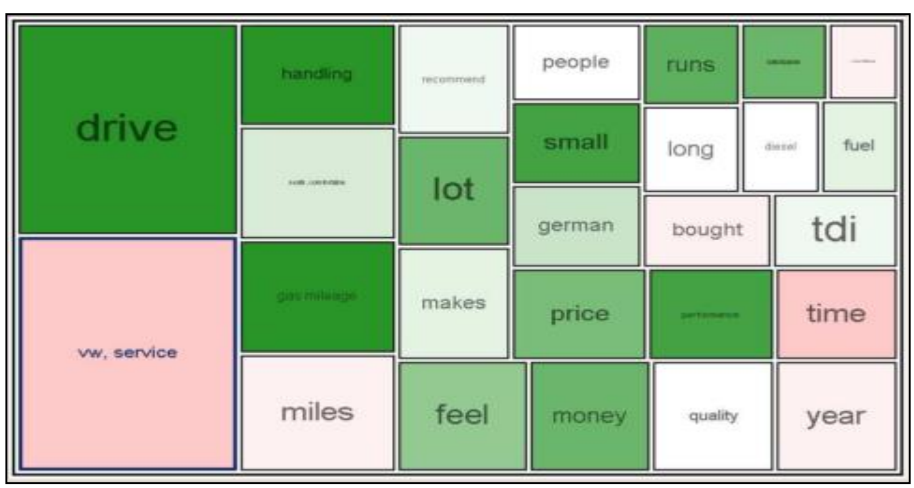


Figure 3.3: Tree Map showing Car's Features and corresponding Sentiment (Gamon et al., 2005)

Similarly, visual summary report provides a quick analysis of printers' reviews (Oelke et al., 2009). It compares the key features of competing products based on their sentiments by adopting a color scale to highlight strengths and weaknesses of competing products.

Boxes represent prominent features and the size of the inner box encodes the percentage of reviews commented on a particular feature as shown in Figure 3.4. The positive and negative trends of opinions on different features were represented by different shades of blue and red colors, respectively. The strength of visual summary is its scalability with respect to the number of features and products. The color of a box in combination with the size of the inner rectangle provides the importance and customers satisfaction/ dissatisfaction about a feature, for instance, the large size of the inner rectangles of the feature print for printer one shows that many customers commented on it, and the dark red color encodes that a relatively large number of customers are satisfied with the print quality of the printer one.

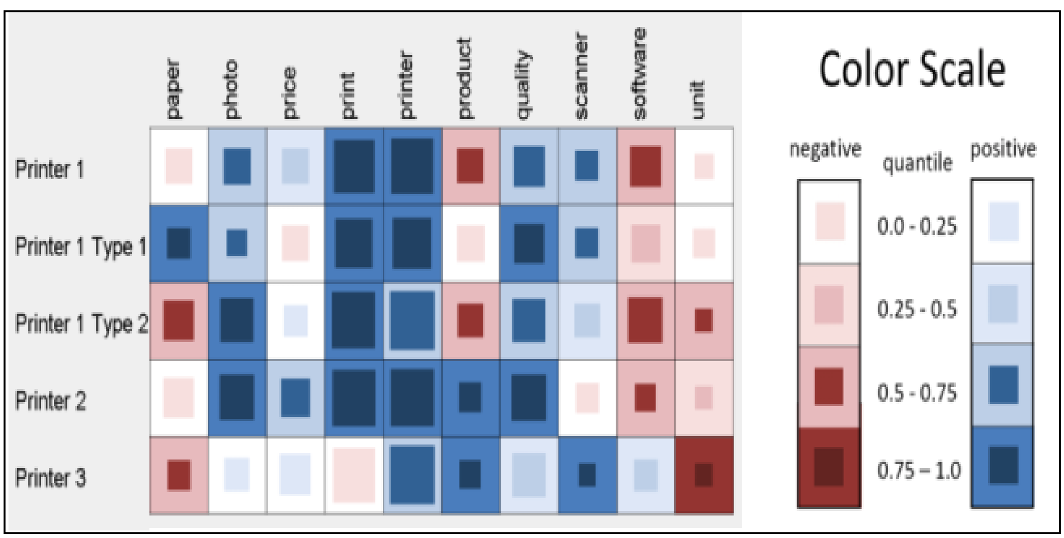


Figure 3.4: Visual Summary represents Printers' Reviews with associated Sentiment (Oelke et al., 2009)

3.4.3. Graphs

Graphs adopted for opinion visualization include coordinated graph (Bjørkelund, Burnett, & Nørvåg, 2012; Chen, Ibekwe-sanjuan, Sanjuan, & Weaver, 2006), line graph (Filho, Brun, & Rondeau, 2012; Miao et al., 2009), and pie chart (Filho et al., 2012; Guo, Zhu, Guo, Zhang, & Su, 2010; Miao et al., 2009; Wang & Araki, 2007). In Chen et al. (2006), the authors visualized the conflicting opinions on the controversial bestseller novel ‘Da Vinci Code’ by coordinated views. Positive and negative terms were represented by the top and the bottom half of the coordinated graph, respectively, as shown in Figure 3.5.

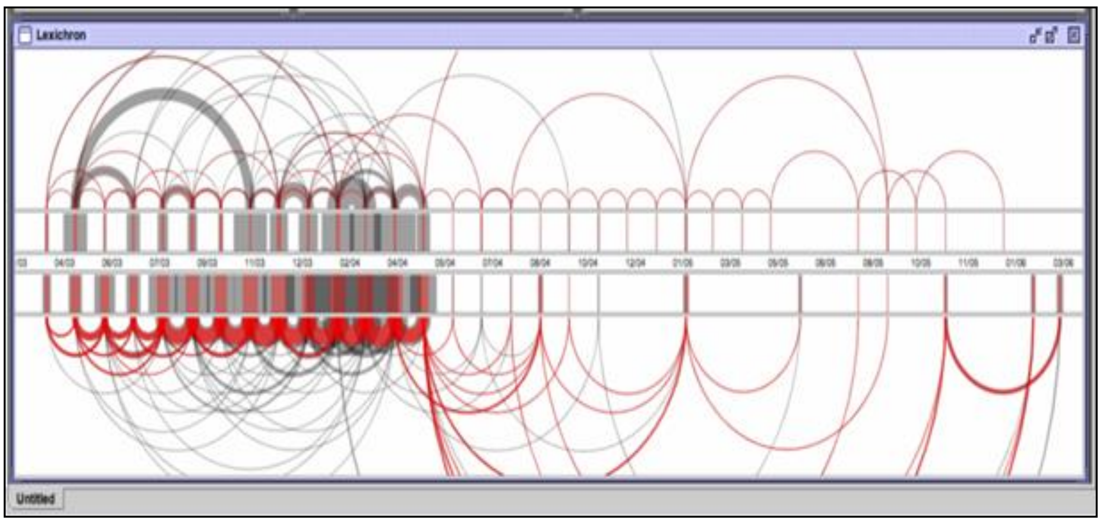


Figure 3.5: Coordinated Graph showing Summary of Positive and Negative Terms (Chen et al., 2006)

Increasing time was encoded from left to right. The thickness of the vertical bar presents the number of terms that appeared in a month. The months in which the common terms appeared were connected by arcs and the thickness of the arcs represent the number of common terms. A spectrum graph was also used to show the dissemination of positive and negative reviews in this work. The strenghts of the graph are the exploration of patterns of term usage and term variation by panning and zooming over time and drill down operation to compare temporal patterns for particular terms.

Similarly, Bjørkelund et al. (2012) applied a graph to present the fluctuations in customers' opinion toward a target hotel on a monthly basis, which depicts sentiment score (average calculated sentiment scores over all reviews within each month), actual score (average score given by authors), sentiment score change (change in the sentiment score), and actual score change (change in the actual score) as shown in Figure 3.6.

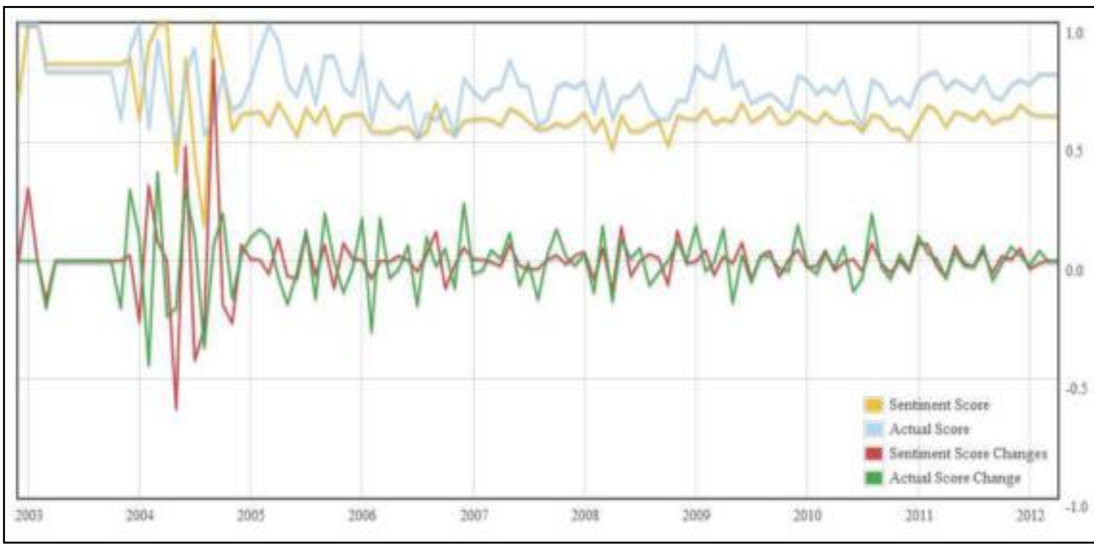
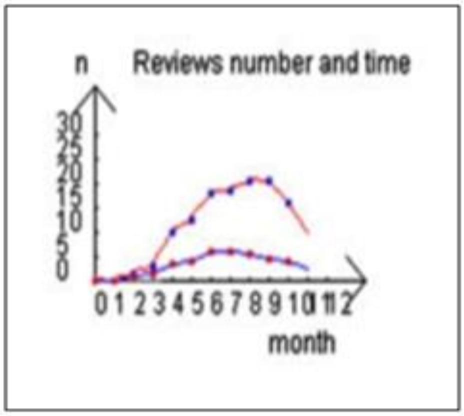


Figure 3.6: Graph representing the Fluctuations in Customers' Opinion (Bjørkelund et al., 2012)

Likewise, a line graph and a pie chart were used to present the number of positive and negative comments about a product over time and the percentage of positive and negative reviews, respectively, in Filho et al. (2012) and Miao et al. (2009) as shown in Figure 3.7. Similarly, a pie chart was deployed in Guo et al., (2010) and Wang & Araki (2007) to show comparison between the number of positive and negative reviews for a target product and competing products (T1, T2, T3, T4, T5), respectively as shown in Figure 3.8. Line graph and pie chart provide comparative analysis in simple, easy to understand, effective way without any pre-knowledge requirement.



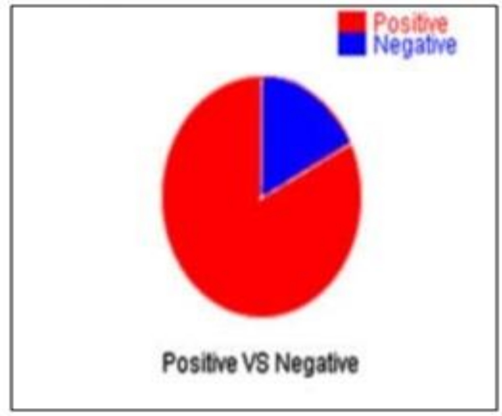


Figure 3.7: Line Graph and Pie Chart showing Opinion Trend Movement and Ratio of Positive and Negative Reviews (Miao et al., 2009)

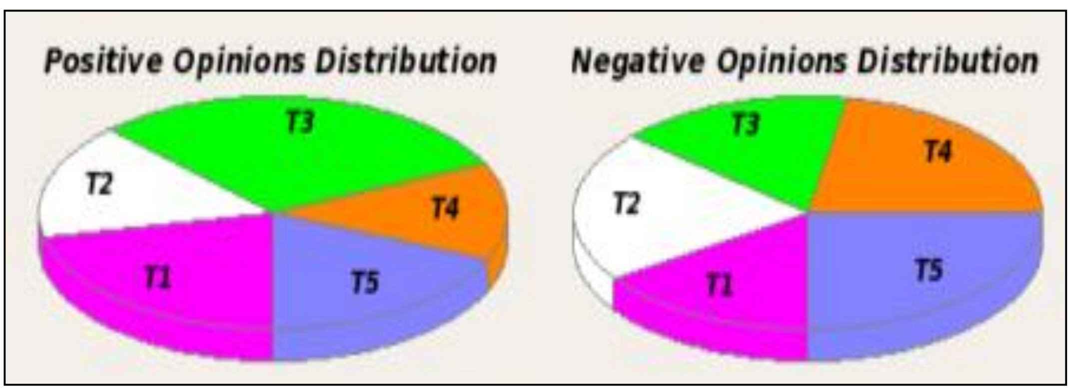


Figure 3.8: Pie Chart showing the Number of Positive and Negative Reviews for Competing Products (Wang & Araki, 2007)

3.4.4. Bar Chart

Bar chart was by various researchers to facilitate visual comparison of feature-based opinions Dey & Haque (2008), Gamon et al. (2008), Guo et al. (2010), Kongthon, Angkawattanawit, Sangkeettrakarn, Palingoon, & Haruechaiyasak (2010), Liu et al. (2005), Wanner et al. (2009) and Wang & Araki (2007) . Wanner et al. (2009) suggested an innovative visual tool to track opinion expressed in RSS news feeds on political parties and their candidates during the US presidential election in 2008. News articles were displayed on a daily basis; two horizontal lines represent one day and each colored object represents one news item as shown in Figure 3.9. Symbols with different colors and shapes were used to show the presence of certain keywords in the news item. The sentiment score of news item was determined by its vertical position. The strengths of this visualization are zooming, filtering, details on demand and similarity search operations, which can be applied in order to highlight interesting trends, particularities, and emotional contents in news items (Wanner et al., 2009).

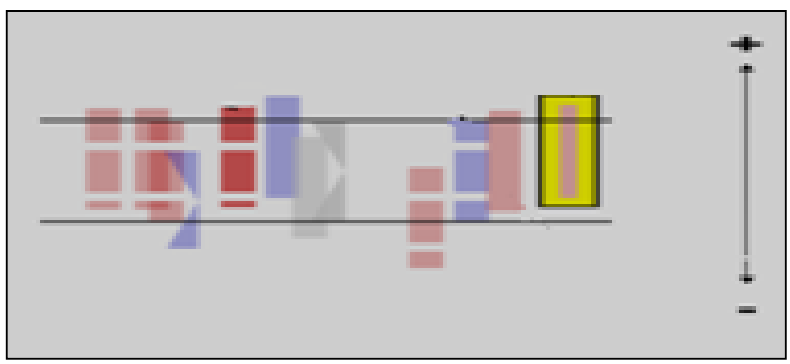


Figure 3.9: Bars with Symbols depicting various Keywords in a News Item (Wanner et al., 2009)

Gamon et al. (2008) deployed a bar chart metaphor glowing bars to present the emotional effect and popularity of political news according to conservative and liberal views as shown in Figure 3.10. The red and blue horizontal bars on each side of a news title depict the number of conservative and liberal references, respectively. News articles are ordered according to their popularity from top to bottom. A glow around the bars portray emotional charge. The glowing bars display the most popular articles, the articles most cited by liberals, the articles most emotionally discussed etc. in a way that is easy to read at a glance.

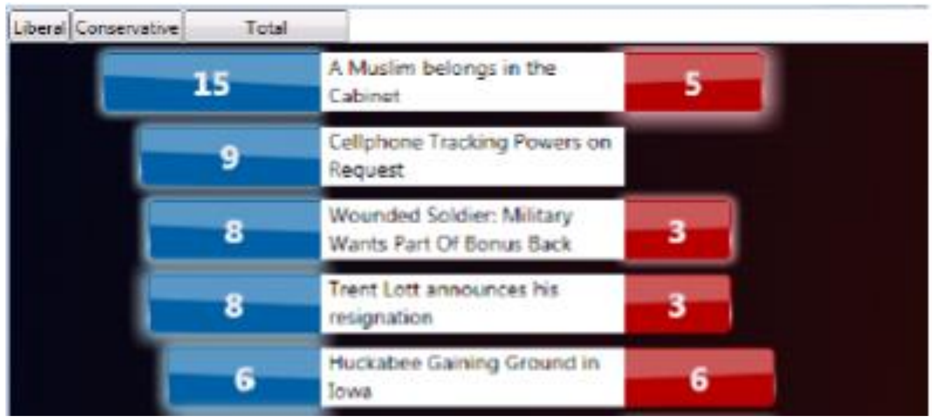


Figure 3.10: Glowing Bars showing Emotional Affect, Popularity and Views about a News Article (Gamon et al., 2008)

In contrast, Liu et al. (2005) utilized vertical bars to highlight the strengths and weaknesses of competitive products. Bars with different colors were used to encode prominent features of competitive products. The number of positive and negative opinion on each feature was represented by the size of bars that lie above and below the x-axis, respectively, as shown in Figure 3.11. Similarly, the comparison between positive and negative opinions on each feature such as ‘Price’ for competing products was displayed using a bar chart in (Wang & Araki, 2007) as shown in Figure 3.12.

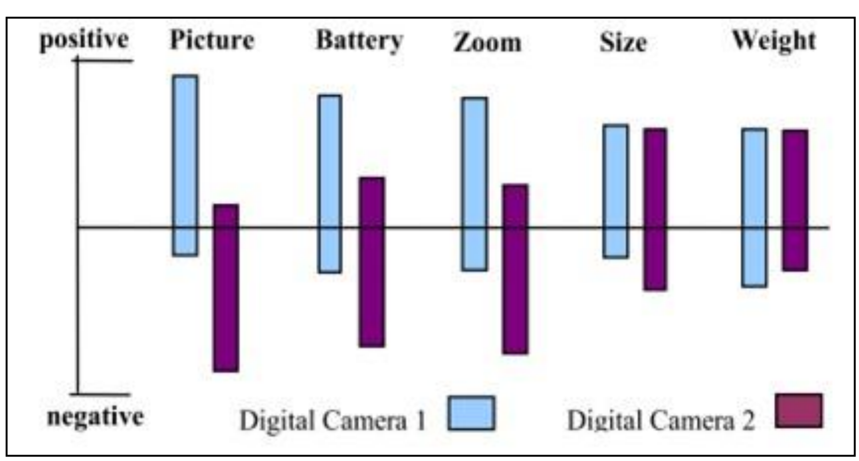


Figure 3.11: Bar Graph Comparing Prominent Features of Competing Products (Liu et al., 2005)

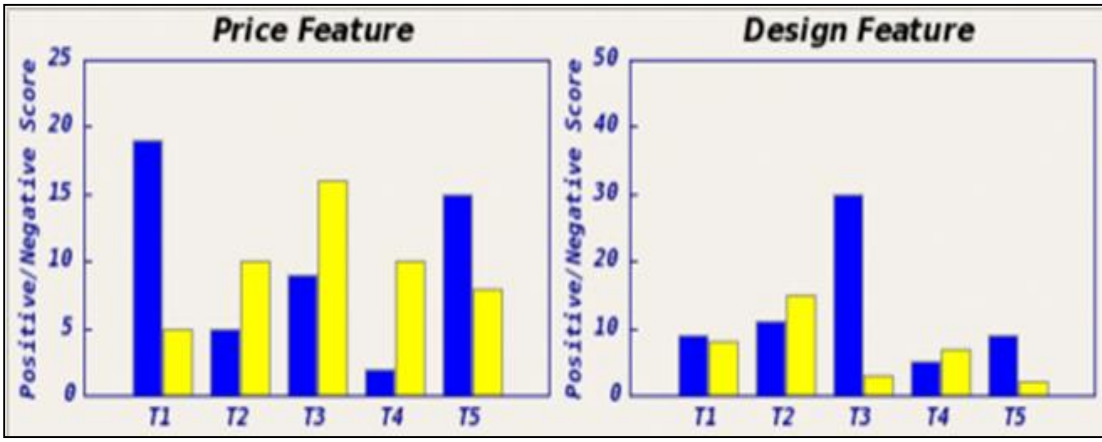


Figure 3.12: Bar Chart Comparing Prominent Features of Competing Products (Wang & Araki, 2007)

In contrast, Dey and Haque (2008) utilized stacked bar charts for comparing different cars' brands on vital features based on the number of positive and negative opinions as shown in Figure 3.13. They also presented prominent features of a brand in terms of positive or negative opinions using a bar chart. Likewise, a stacked bar chart was utilized in (Kongthon et al., 2011) to depict the distribution of sentiment on different features of a target hotel.

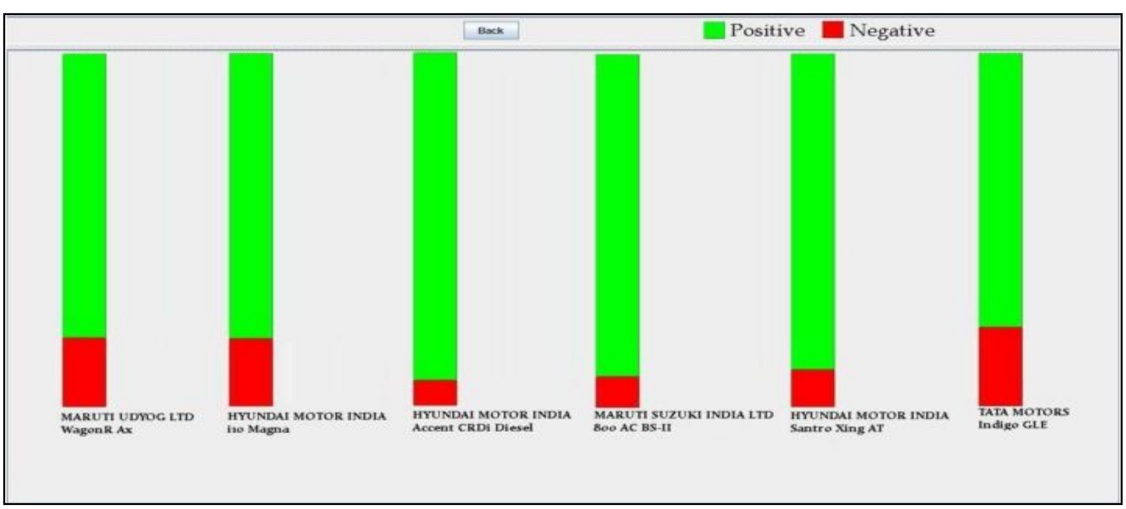


Figure 3.13: Stacked Bar Chart Comparing Cars' Brands on Vital Features based on the number of positive and negative opinions (Dey & Haque, 2008)

Similarly, Guo et al. (2010) exploited a stacked bar chart to compare positive or negative opinions on prominent product features. Traditional bar charts were used in these studies to visualize the distribution of positive and negative opinions that exist within a review document and are much more scalable with respect to the number of features and the number of products that are displayed.

3.4.5. Maps

The final visualization techniques involve the use of map, i.e. positioning map (Morinaga, Yamanishi, Tateishi, & Fukushima, 2002a), comparative relation map (Xu et al., 2011), Google map (Bjørkelund et al., 2012), pixel map calendar and time density plot (Rohrdantz et al., 2012b), pixel cell-based sentiment calendar and geo map (Hao et al., 2011), and key term geo map (Hao et al., 2013).

Morinaga et al. (2002) utilized a positioning map to compare competitive cellular phones based on four characteristics (features), which were plotted around each cellular phone as shown in Figure 3.14. For instance, Cellular Phone A has the best reputation with no problem(s), fast, future and benchmark results as characteristic words. It can be concluded that Cellular Phone A has good reputation ('no problem(s)', 'fast', 'benchmark results') whereas Cellular Phone C has bad reputation ('doesn't work', 'slow') from Figure 3.14. The strength of the positioning map is the simplicity, however, it has scalability problem with respect to the number of features and products.

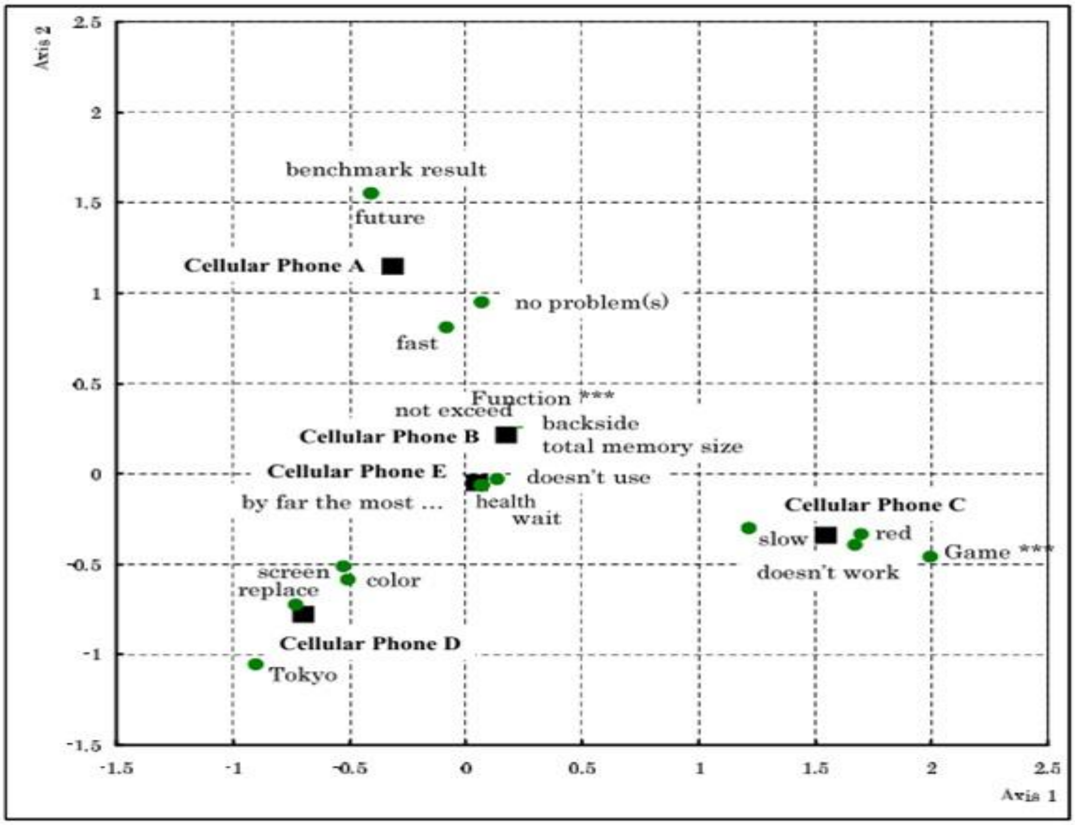


Figure 3.14: Positioning Map representing Competing Cellular Phones with their Characteristics (Morinaga et al., 2002)

Likewise, Xu et al. (2011) introduced comparative relation map to facility decision making process. The comparative relation map displays comparative relations between competing products from customer reviews. For instance, Nokia E71 comparative relations with its competitors, i.e. Blackberry Curve, Blackberry Bold 9000, Nokia E61, and iPhone based on key features, such as function, screen, looks, size etc. are depicted in Figure 3.15.

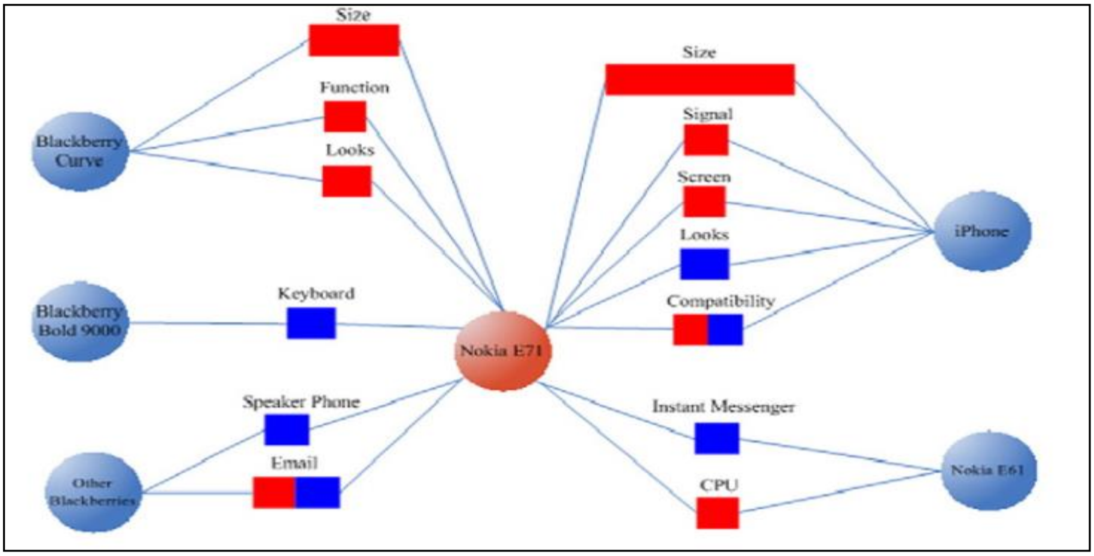


Figure 3.15: Comparative Relation Map Comparing Competing Mobile Phones (Xu et al., 2011)

The red and blue boxes show the number of comparative relationships for Nokia E71 and the number of comparative relationships for competitive products, respectively (Figure 3.15). The comparative relation map was very helpful to (i) highlight the relative strengths and weakness of products, (ii) analyze threats from competitors and enterprise risks, (iii) support decision making and risk management, and (iv) design new products and marketing strategies.

In Bjørkelund et al. (2012), a Google map was deployed to facilitate users to identify good hotels and good areas to stay in as shown in Figure 3.16. The average sentiment and the number of reviews for each hotel were presented as a colored circle of varying size; green for positive and red for negative. The visualization provides a user-friendly view of good and bad geographical hotel areas based on customers' opinions.

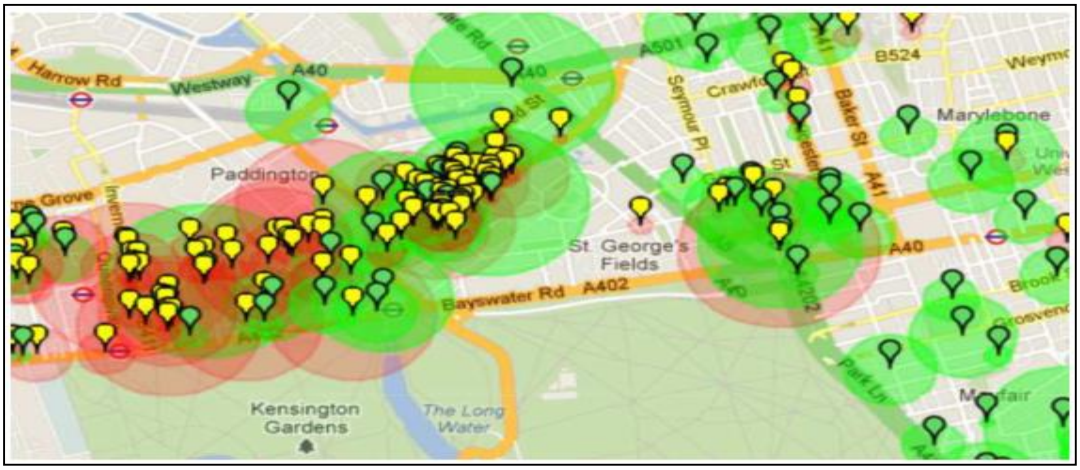


Figure 3.16: Google Map showing Good and Bad Hotels (Bjørkelund et al., 2012)

Pixel map calendar and time density plot were utilized in Rohrdantz et al. (2012b) for visual text time series analysis. Each document represented by one pixel and the color of the pixel indicates the average sentiment of the document; green for positive, yellow for neutral, and red for negative sentiments (Figure 3.17). The X-axis encodes days and the Y-axis corresponds to months with years. An enlarged view of April and May 2008 is presented in Figure 3.17, which shows the overall sentiment and the sentiment on feature ‘password’. It can be deduced that the ‘password’ is infrequent feature and occurs mostly in negative contexts.

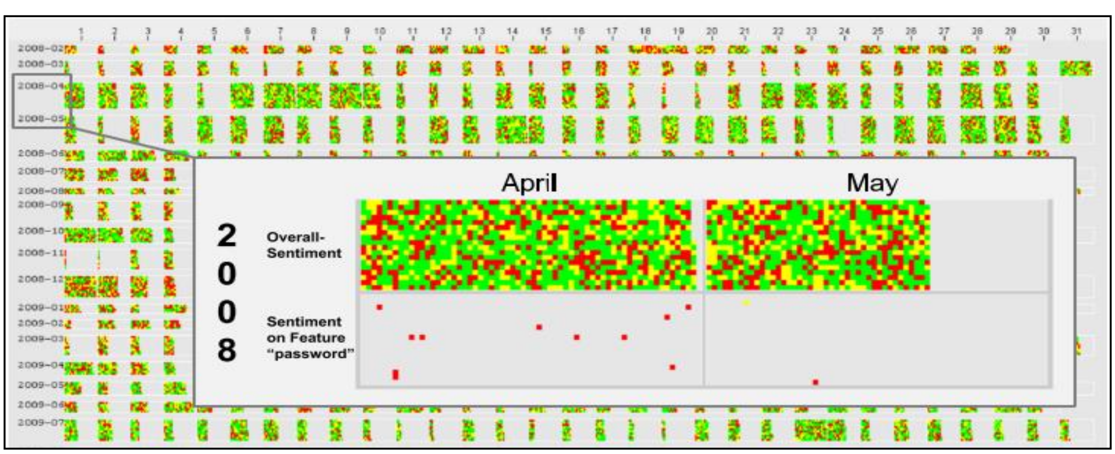


Figure 3.17: Pixel Map Calendar showing the Visual Analysis of Text Stream (Rohrdantz et al., 2012b)

Additionally, for each feature one individual time density plot was also created. The time density plot consists of a sequential sentiment track and a time density track. The sequential sentiment track represents all documents discussing a feature in a sequential order. Each colored bar corresponds one document that contains the feature. The height and color of a bar reflect the certainty level and polarity of the feature, respectively as shown in Figure 3.18. The pixel map calendar is a suitable visualization that provides an overview of visual analysis of text streams and the time density plot provides detailed insight by further data exploration of time interval patterns for multiple features.

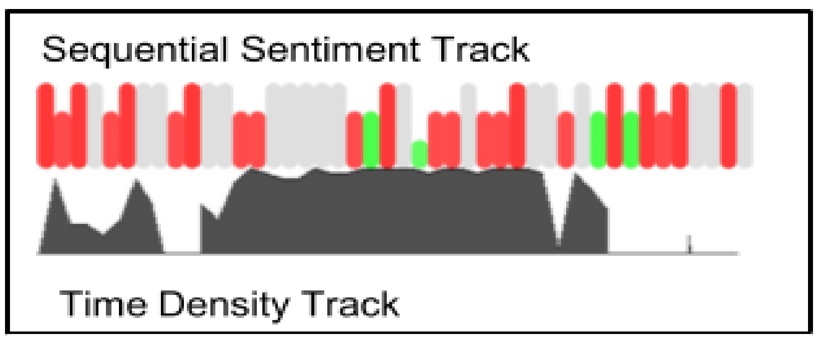


Figure 3.18: Time Density Plot showing the Visual Analysis of Features (Rohrdantz et al., 2012b)

Hao et al. (2011) and Hao et al. (2013) visualized the high-volume Twitter data using pixel cell-based sentiment calendar and geo map that highlight interesting tweets based on their density, semantic, and influential characteristics. Figure 3.19 shows a pixel cell-based sentiment calendar in which rows encode topics and columns present time interval (hours). A cell reflects an opinion. The color of a cell shows the sentiment value; green for positive, gray for neutral, and red for negative. The pixel cell-based sentiment calendar shown in Figure 3.19 provides the analysis of Twitter comments on the movie Kung-Fu Panda. Geo map was also introduced in this work to display the geographical distribution of tweets as shown Figure 3.20. Each point represents a tweet. The color of a point reflects the sentiment value by utilizing same color scheme used in the pixel cell-based sentiment calendar. The visualizations provide visual analysis of Twitter data stream with respect to users' comments to visualize large volumes of data in a single view enabling users to comprehend the distribution of customers' feedback on specific features over time and geographic locations. The advantage of the sentiment calendar is its scalability with respect to both the number of comments and the number of features.

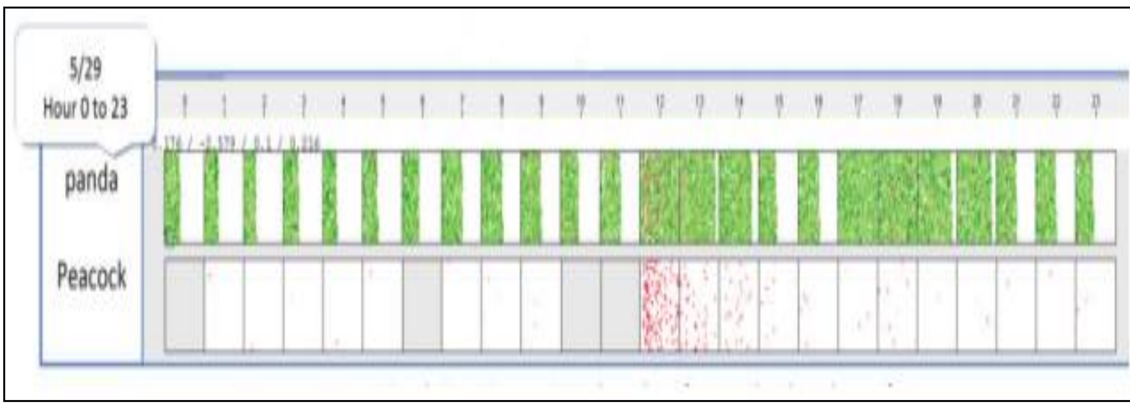


Figure 3.19: Pixel Cell-based Sentiment Calendar showing the Analysis of Comments on the movie Kung-Fu Panda on Twitter (Hao et al., 2011)

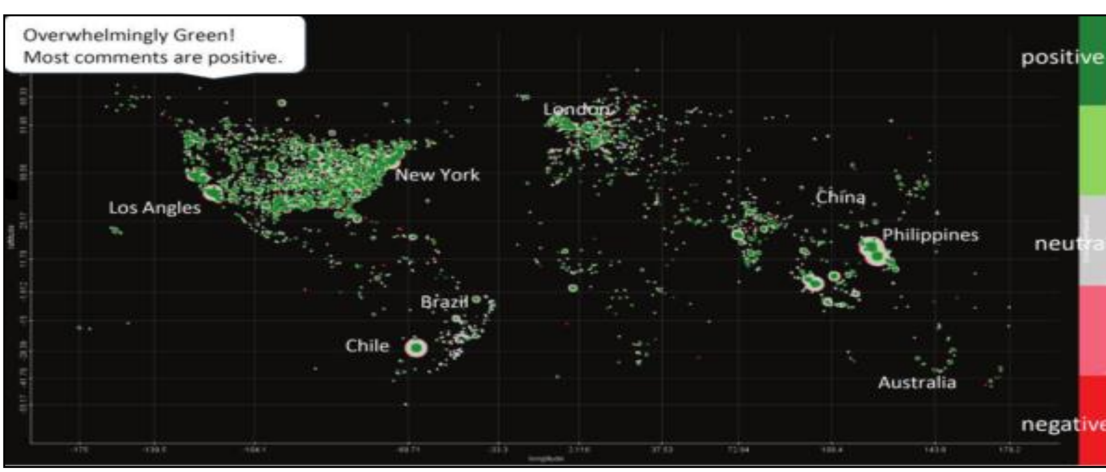


Figure 3.20: Geo Map showing the Distribution of Tweets on the Movie Kung-Fu Panda (Hao et al., 2011)

In another study, Hao et al. (2013) introduced a key term geo map to identify significant terms in a geographical location. In the map, the font size encodes the significant value indicating the level of importance of that particular term and the color represents the sentiment value representing the average opinion (red for negative, green for positive and grey for neutral) as shown in Figure 3.21. The visualization enables analysts to examine the significance and the sentiment distribution of key terms over different geographic locations.

Figure 3.21: Key Term Geo Map showing Significant Key Terms (Hao et al., 2013)

3.5 Research Issues and Challenges

Customers' word-of-mouth on the Web is an enormous source of decision-oriented information, which presents many challenges and opportunities to the opinion mining research community. Current opinion mining techniques are still primitive in nature and pose many challenges, issues, and opportunities for researchers. This section highlights some of the common research issues and challenges in the field of opinion mining.

The fundamental research issue and challenge in opinion mining is to distinguish between objective and subjective information in order to identify which document or part of the document contains opinionated text (Khan, Baharudin, Khan, & Malik, 2009; Pang & Lee, 2008). There are two approaches for subjectivity classification, supervised and un-supervised learning. Most of the existing subjectivity classification approaches are supervised and are extensively studied in the literature (Barbosa & Feng, 2010; Benamara & Popescu, 2011; Pang & Lee, 2004; Raaijmakers & Kraaij, 2008; Wiebe et al., 1999; Wiebe & Riloff, 2005; Wilson et al., 2004; Yu & Hatzivassiloglou, 2008), whereas unsupervised subjectivity classification has only been studied by Wiebe (2000).

The second issue is the extraction of features. The review of the literature revealed that four techniques are used to solve the feature extraction problems; namely, frequent nouns and noun phrases based extraction (Blair-goldensohn et al., 2008; Hu & Liu, 2004; Ku et al., 2006; Long, Zhang, & Zhut, 2010; Popescu & Etzioni, 2005; Scaffidi et al., 2007; Zhu et al., 2009), opinion and target relations based extraction (Blair-goldensohn et al., 2008; Hu & Liu, 2004; Qiu et al., 2011; Somasundaran & Wiebe, 2009; Wu, Tan, & Cheng, 2009; Zhuang et al., 2006), supervised learning based extraction (Choi & Cardie, 2010; Jakob & Gurevych, 2010a; Jin & Ho, 2009; Kobayashi, Inui, & Matsumoto, 2007; Li, Connexis, Liu, & Street, 2010; Liu et al., 2005) and topic modelling based extraction (Branavan, Chen, Eisenstein, & Barzilay, 2008; Li et al., 2010; Lin, Road, & Ex, 2009; Lu et al., 2009; Mei et al., 2007; Titov & McDonald, 2008).

The third issue is the classification of opinions as positive, negative, and neutral. Like the subjectivity classification, most existing opinion classification techniques use supervised (Abbasi, Chen, & Salem, 2008; Cui, 2006; Dave et al., 2003; Gamon, 2004; Li et al., 2010; Mullen & Collier, 2004; Ng, Dasgupta, & Arifin, 2006; Paltoglou & Thelwall, 2010; Pang et al., 2002; Pang & Lee, 2004; Xia & Zong, 2011), or unsupervised learning (Ding et al., 2008; Hu & Liu, 2004; Kim, Rey, & Hovy, 2004; Taboada, Brooke, & Voll, 2011; Turney, 2002). Very few studies have been conducted to classify opinions using semi-supervised learning (Dasgupta & Ng, 2009; Li, Wang, Zhou, Yat, & Lee, 2011; Zhou, Chen, & Wang, 2010).

According to Etzioni et al. (2005), opinions vary in terms of their strength. Thus, the fourth issue is to identify the strength of an opinion, which, currently, has received little attention in the literature. The fifth issue is how to summarize the results in effective ways because the opinion summary is different from a traditional single document or multi-document summary (Liu, 2012). Some of the works done addressing this issue includes (Blair-Goldensohn et al., 2008; Carenini, Ng, & Pauls, 2006; Hu & Liu, 2004; Ku et al., 2006; Liu et al., 2005; Lu et al., 2010; Tata & Eugenio, 2010; Zhuang et al., 2006). The sixth issue is how to summarize the conflicting opinions (Carenini et al., 2006). The seventh issue is an effective visualization of the opinion data because careful visualization designs are required to present the opinions (Wu et al., 2010). Currently, opinion mining tools are used by data analysts. Therefore, these tools require better usability and user friendliness when used by customers (Osimo & Mureddu, 2012). Considerably less work has been done on opinion visualization (Gregory et al., 2006; Hao et al., 2013; Oelke et al., 2009; Rohrdantz et al., 2012b; Wanner et al., 2009; Wu et al., 2010; Xu et al., 2011).

The eighth challenging issue is the detection of spam, fake reviews, outliers, and the reputation of the reviewers (Pang & Lee, 2008). Some authors focused on supervised spam detection (Jindal & Liu, 2008; Li et al., 2011; Ott, Choi, Cardie, & Hancock, 2011) while others focused on un-supervised spam detection (Jindal, Morgan, & Liu, 2010; Lim, Nguyen, Jindal, Liu, & Lauw, 2010; Wang, Xie, Liu, & Yu, 2012). Online reviews suffer from non-expert opinions because they are normally written by laymen (Khan et al., 2009). Another issue is to extract comparative opinions on competitive products, which involve multiple entities and the identification of the relationships between these entities (Xu et al., 2011). This issue was not extensively studied in the literature except for a few (Fiszman, Demner-fushman, Lang, Goetz, & Rindflesch, 2007; Jindal & Liu, 2006; Xu et al., 2011; Yang, 2011).

Mostly, opinion mining techniques are domain dependent because customers express opinions about a specific issue, product, or topic and it is hard to generalize these techniques (Khan et al., 2009). Therefore, the development of domain independent techniques is another challenging research issue. A semantic classifier trained for one domain often performs poorly if applied to another domain (Liu, 2012). Some researchers addressed this issue by using labelled training data for a different domain (Aue & Gamon, 2005; Yang, Si, & Callan, 2006), while others addressed it without any labelled training data (Andreevskaia & Bergler, 2008; Blitzer, 2006; Gao & Li, 2011; Pan, Ni, Sun, Yang, & Chen, 2010; Tan, Wu, Tang, & Cheng, 2007). Other issues include the identification of authority, authenticity and credibility of opinion source (Conrad & Leidner, 2008; Seerat & Azam, 2012) and the monitoring of the customers' opinions trend movement (Chauchat, Eric, Lumière, & Mendès-france, 2008), which is addressed by plotting the number of positive and negative comments over time using a line graph (Miao et al., 2009).

The nature of opinion data presents a great challenge. Large-scale, heterogeneous, informally written, free text, high-dimensional and highly diverse opinion data present additional research issues and challenges in mining, summarization, and visualization of opinion data (Khan et al., 2009; Pang & Lee, 2008; Wu et al., 2010). Reviews are generally written by people who are different in terms of their knowledge, writing style, and the use of abbreviations all of which present further issues (Khan et al., 2009).

In the last decade, product feature scoring based on semantic polarities has received considerable attention from researchers and become a new research direction in the opinion mining domain. Representative work on feature scoring includes (Eirinaki et al., 2012; Kunpeng et al., 2010; Lei et al., 2010; Li et al., 2011; Moghaddam & Ester, 2010; Scaffidi et al., 2007; Yang et al., 2010).

Online reviews are growing at a fast pace and vary greatly in quality, thus it becomes difficult to identify high quality helpful reviews. Most of existing opinion mining systems ignore the quality of the reviews; therefore, effective review quality evaluation methods are required to identify high quality reviews (Chen & Tseng, 2010). Existing review quality approaches are based on the textual and/or social features (statistics) of the reviews. Existing works (Chen & Tseng, 2010; Ghose & Ipeirotis, 2007, 2011; Kim et al., 2006; O'Mahony & Smyth, 2009; Zhang & Varadarajan, 2006) are based on textual features, i.e. number of adjective and nouns, length of the review etc. in conjunction with social features.

This section provided an extensive review of existing research issues and challenges in the field of opinion mining, including subjectivity analysis, feature and opinion extraction, opinion classification and strength identification, opinion summarization and visualization, detection of spam reviews, identification of comparative sentences, reviews and features ranking. The current study focuses (I) reviews ranking, (ii) feature ranking, and (iii) opinion visualization.

Chapter four describes the research mythology used in this research work to develop an opinion mining system based on novel reviews ranking and opinion visualization techniques according to the users' perspective, and a feature ranking method that utilizes opinion strength.

Chapter 4 : Methodology

In this research, a prototype for an opinion mining system called Opinion Analyzer is implemented that assimilates high quality review in the proposed feature ranking approach and provides an opinion-strength-based feature-level summary and feature ranking. The Opinion Analyzer takes a collection of reviews as input, and outputs a set of extracted features along with their corresponding ranking and an opinion-strength-based summary. Further, the Opinion Analyzer provides feature ranking according to the positive, negative and overall ranks of prominent features of a target product. Moreover, an opinion summary is presented by the system that highlights the different levels of opinion strength associated with the features.

This chapter describes the architecture of the Opinion Analyzer that incorporates all the processes from review ranking, feature ranking to present results to end users. Experimental data set and setup used to measure the effectiveness of the Opinion Analyzer are also discussed in this chapter. Moreover, methodology used for accessing the users'preferences about existing opinion visualizations and usability of the opinion-strength-based visualization are presented in this chapter.

4.1. Opinion Analyzer

In the Opinion Analyzer, each review is represented by a tuple of two elements [MD, B]. The MD describes the metadata of a review and consists of H and R, where H represents the helpfulness ratio (helpfulness votes/total votes*100) of the review and R reflects the users' rating. The B reflects the body of a review and contains a set of sentences in the body of reviews. The users' rating is added to the review tuple (Miao et al., 2009) for review ranking. Similarly, review sentences are represented by a tuple, that is an extension of the tuple discussed in Liu (2006) and Miao et al. (2009). The proposed sentence tuple consists of four elements [F, P, St, C], where F represents a product feature, P reflects the semantic polarity of the feature F in the same sentence, St defines the opinion strength and C describes the content of the sentence. The proposed tuples are presented below:

$$Review = \{(MD, B)\}$$

$$MD = \{H, R\}$$

$$B = \{S1, S2, S3, \ldots., Sn\}$$

$$\forall\, S \in B, \qquad S = \{F, \mathrm{P}, \mathrm{St}, \mathrm{C}\}$$

Figure 4.1 shows the graphical presentation of the proposed review tuple.

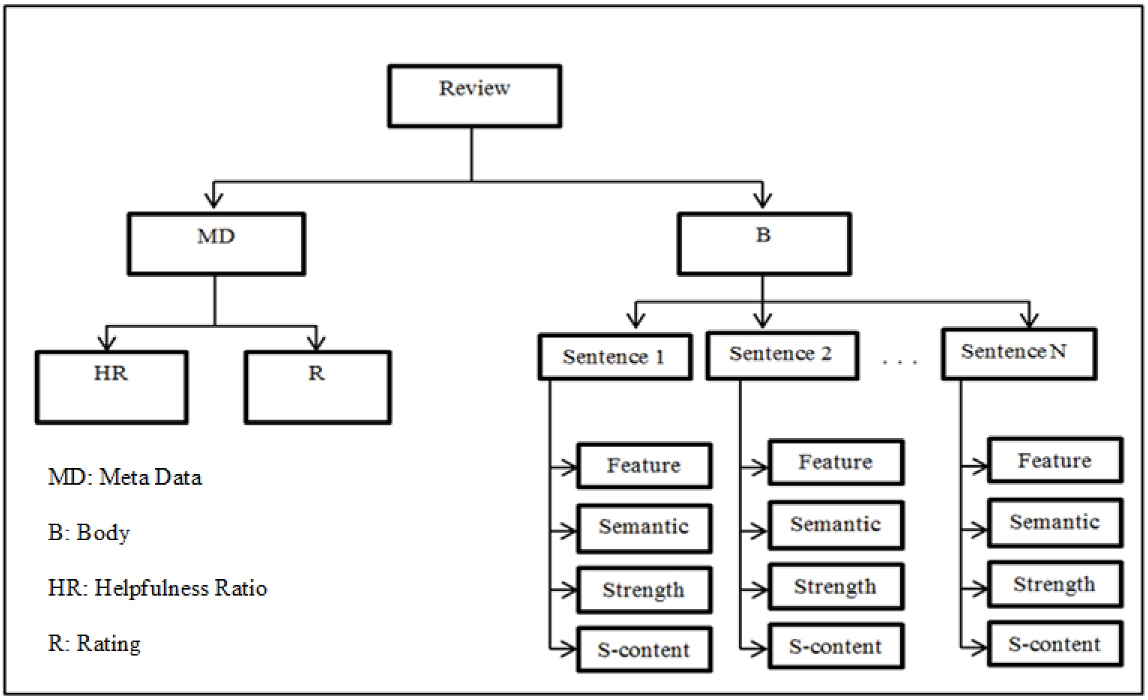


Figure 4.1: Proposed Review Tuple

Similar to the proposed review tuple, a feature tuple is defined that includes three elements [W, POS, NEG], where W is the feature weight (feature frequency), POS and NEG are the accumulated strengths of positive and negative opinions associated with the feature, respectively. Opinion strength is estimated in the range from +1 to +3 for positive orientation and -3 to -1 for negative orientation (one for weakest and three for strongest), for instance, +3 is assigned to excellent and +1 is assigned to good, (similarly, -3 is assigned to terrible and -1 is assigned to bad) in this work.

Similarly, the POS and NEG are further represented by tuples based on the definition of opinion strength defined in Binali et al. (2009) and Osimo & Mureddu (2012). The tuple for POS is [WP, MP, SP] and the tuple for NEG is [WN, MN, SN] as shown below in Figure 4.2, where WP is the weakly positive opinion with opinion strength +1, MP is the mildly positive opinion with opinion strength +2, SP is the strongly positive opinion with opinion strength +3, WN is the weakly negative opinion with opinion strength -1, MN is the mildly negative opinion with opinion strength -2 and SN is the strongly negative opinion with opinion strength -3.

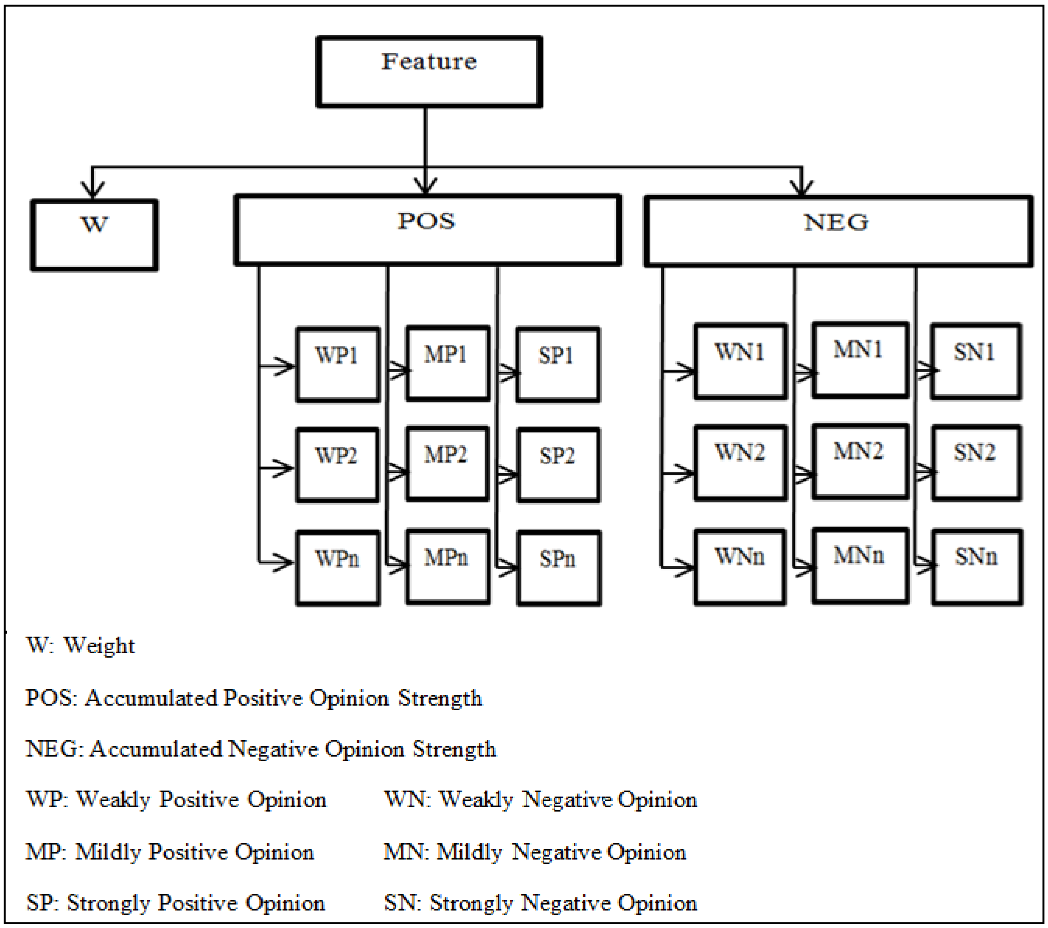


Figure 4.2: Proposed Feature Tuple

The proposed feature tuple is mathematically shown below:

$$Feature = \{W, POS, NEG\}$$

$$POS = \{WP_1, WP_2 \ \dots, WP_i, \quad MP_1, MP_2, \dots\dots, MP_j, \quad SP_1, SP_2, \dots\dots, SP_k\}$$

$$NEG = \{WN_1, WN_2 \ \dots, WN_m, \quad MN_1, MN_2, \dots\dots, MN_n, \quad SN_1, SN_2, \dots\dots, SN_o\}$$

Where i represents the number of weakly positive opinion words, j reflects the number of mildly postive opinion words and k describes the number of strongly positive opinion words associated with the feature. Similarly, m presents the number of weakly negative opinon words, n encodes the number of mildly negative opinion words and o defines the number of strongly negative opinion words connected with the feature.

Consider the following review shown in Figure 4.3.

Figure 4.3: A Sample Review

The above mentioned review is expressing opinions on '*picture quality*', '*battery*', and '*viewfinder*' features of a camera. In the first sentence of the review, the feature '*picture quality*' is described by the opinion word '*good*', the semantic orientation of the opinion word '*good*' is '*positive*', and the corresponding strength of the opinion word '*good*' is +1 (Weakly Positive). The opinion word '*poor*' is associated with the feature '*battery*' in the second sentence of the review, the semantic orientation of the opinion word '*poor*' is '*negative*' and the strength of the opinion word '*poor*' is -1 (Weakly Negative). In the last sentence of the review, the feature '*viewfinder*' is connected with the opinion word '*very good*', the semantic orientation of the opinion word '*very good*' is '*positive*' and corresponding opinion strength is +3 (Strongly Positive) in this case. The resulting tuple of the review shown in Figure 4.3 is illustrated in Figure 4.4.

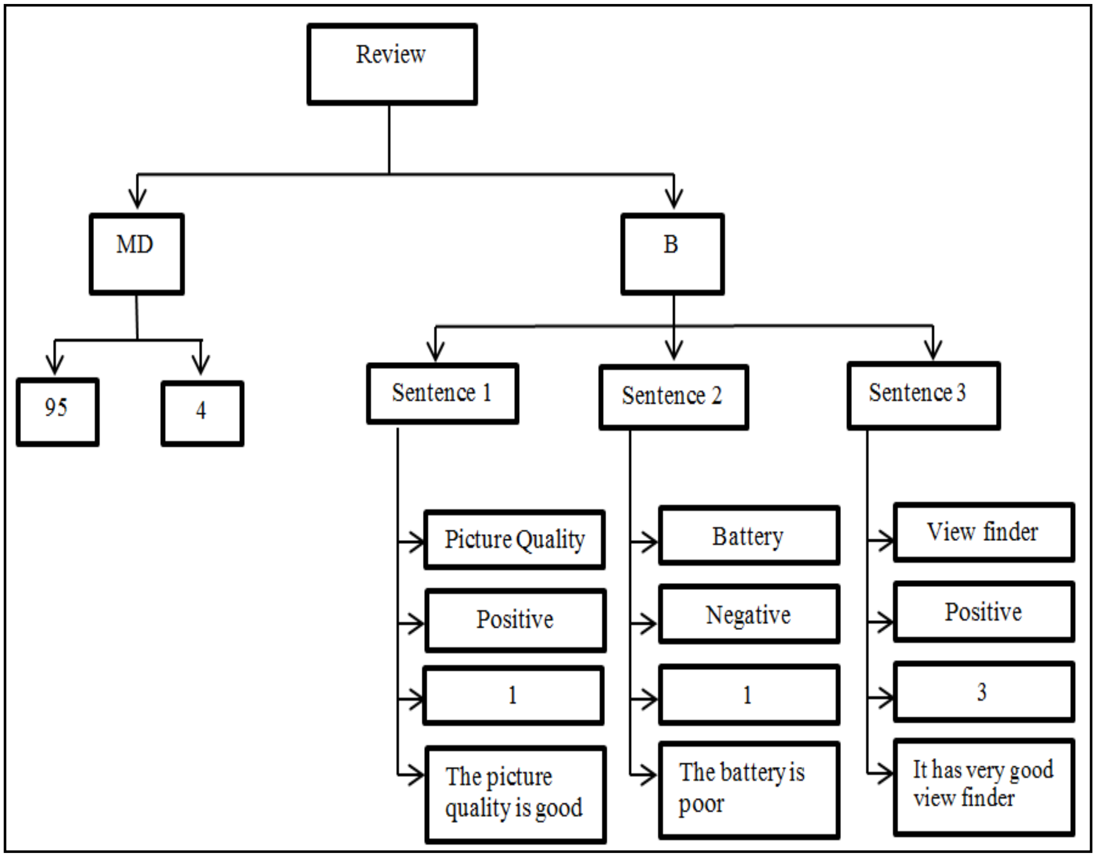


Figure 4.4: An Example of Review Tuple

Consider the following reviews shown in Figure 4.5.

Figure 4.5: Sample Reviews

The reviews are expressing opinions on features '*picture quality*', '*battery*', and '*view finder*' of a target camera. The corresponding feature tuple of the feature '*battery*' is shown in Figure 4.6. The weight of the feature '*battery*' is three as it was discussed three times in all of the reviews. The feature '*battery*' is described by the opinion words '*good'* , *' poor'*, and *'disappointing*' and the corresponding strengths of the opinion words are +1 (Weakly Positive), -1 (Weakly Negative), and -2 (Mildly Negative), respectively. Therefore, the POS of the feature battery is +1 as only one positive opinion word (Weakly Positive) is connected with the feature having an opinion strength +1, whereas NEG of the feature is -3 (-2 + -1) as two negative opinion words (Weakly Negative and Mildly Negative) are associated with the feature having -1 and -2 opinion strengths (Figure 4.6).

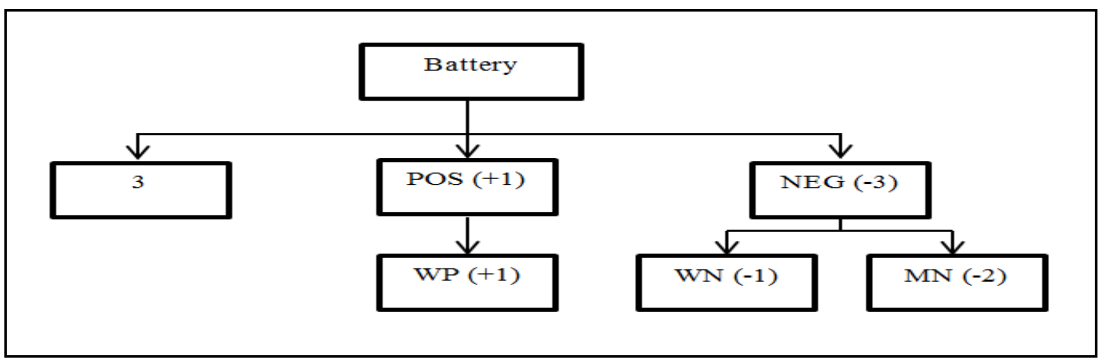


Figure 4.6: An Example of Feature Tuple

The main components of the Opinion Analyzer are shown in Figure 4.7. It consists of five components: (i) data pre-processor, (ii) feature and opinion extractor, (iii) review ranker, (iv) feature ranker, and (v) opinion visualizer. Inputs of the Opinion Analyzer are review document and users' preferences, and its output is an opinion summary.

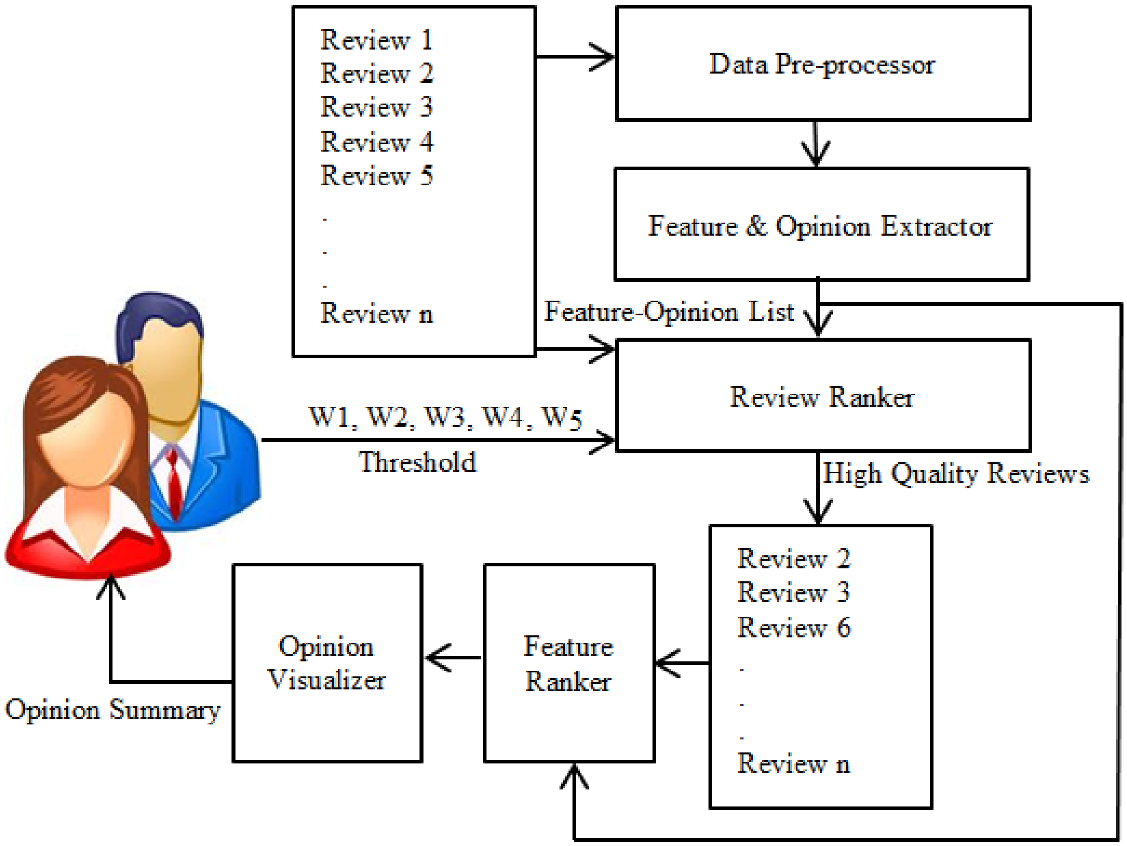


Figure 4.7: Architecture of Opinion Analyzer

In the following section, we briefly outline the main components of the Opinion Analyzer, as illustrated in Figure 4.7.

4.1.1 Data Pre-Processor

Data pre-processor is responsible for cleaning and preparing review document for opinion mining as online reviews contain uninformative parts such as HTML tags, scripts and advertisements. The aim of the data pre-processor is the pre-structuring of reviews' text in order to prepare it for opinion mining by transforming it into a processible structure using methods from the fields of natural language processing and information retrieval.

The data from the review document are pre-processed to convert the data in the format which is acceptable to the Opinion Analyzer. In this thesis, a real data set of amazon.com utilized by Liu and Hu (2004) is used for experimental purpose. The data set is pre-processed and publically available. In the data set of Liu and Hu (2004), the tag [t] indicates that the sentence following [t] is the title of the review (Figure 4.8). Further, a new line in the review is expressed with '##' symbol. Features '*camera*', '*picture*', '*color*', '*white balance*', and '*optical zoom*' of Canon PowerShot G3 camera are marked with their corresponding opinion strength and semantic orientation in the data file, for instance, the feature '*picture*' is marked with a positive opinion intensity of two as highlighted in Figure 4.8. However, both the helpfulness ratio and rating are not available in the data set. Therefore, these were calculated and added in the current study. Moreover, feature and opinions are not marked in the title of reviews. These were marked in the study.

```
[t]great gadget
##i bought this last week through amazon .
##got a great deal from a reputable seller .
camera[+2]##i love this camera .
##i am still trying to figure out the may options it has .
picture[+2]##took hundreds of pictures and they were great .
color[+2],picture[+2],white balance[+2]##great colors , pictures and white balance .
optical zoom[+1]##has 4x optical zoom which is higher than any other in the same price range .
##it is generally overpriced a little bit but you get what you are paying for .
```

Figure 4.8: A sample review of Canon G3 Camera from Data Set

The helpfulness ratio of a review was calculated from the helpfulness votes of a review (provided by amazon.com) for each review in the file, for instance, the helpfulness ratio of the review shown in Figure 4.9 is 75 (3/4*100) as three people out of four found the review helpful. Two tags, namely [h] and [r] were introduced to represent helpfulness ratio and users' rating, respectively. Figure 4.10 shows the final review after the helpfulness ratio and rating were included, and features and opinions in the title were marked, in which [h][75] represents the helpfulness ratio (75%) and [r][5] describes the rating of the review, that is, 5 star.

3 of 4 people found the following review helpful

Great gadget, June 2, 2003

By daefash (Phoenix, AZ) - See all my reviews

Verified Purchase (What's this?)

This review is from: **Canon PowerShot G3 4MP Digital Camera w/ 4x Optical Zoom (Electronics)**

i bought this last week through amazon.. got a great deal from a reputable seller. i love this camera. i am still trying to figure out the may options it has. took hundreds of pictures and they were great. great colors, pictures and white balance. has 4x optical zoom which is higher than any other in the same price range. it is generally overpriced a little bit but you get what you are paying for.

Figure 4.9: A Sample Review after adding Helpfulness Ratio and Rating

[h][75][r][5][t] gadget [+2] great gadget
##i bought this last week through amazon .
##got a great deal from a reputable seller .
camera[+2]##i love this camera .
##i am still trying to figure out the may options it has .
picture[+2]##took hundreds of pictures and they were great .
color[+2],picture[+2],white balance[+2]##great colors , pictures and white balance .
optical zoom[+1]##has 4x optical zoom which is higher than any other in the same price range .
##it is generally overpriced a little bit but you get what you are paying for .

Figure 4.10: A sample Review of Canon PowerShot G3 Camera from Amazon.com

The data pre-processor performs different pre-processing steps, namely, tokenization, stop word filter, conversion of reviews' text to lower case, removal of non-alphabetic characters, word stemming, spell checker and POS tagging on the processed data set. At first, the text of the review document is converted into lower case. Secondly, a stop word filter is applied to eliminate stop words from the document according to a given stop word list. Thirdly, the inflectional and derivationally related forms of words are converted into a base form using word stemming. Then spell checking is performed to remove noise from the document.

Lastly, POS tagging is applied to assign a part of speech category to each word in the document. For instance, consider the following sentence:

Sentence 11: ‘*My wife says, the canon cameras are easier to use*’.

The tokenization removes the punctuation from the sentence and returns a list of words. The stop word filter removes the words ‘*my*’, ‘*the*’ and ‘*to*’ from the sentence. The word stemming reduces the words ‘*says*’ and ‘*easier*’ into ‘*say*’ and ‘*easy*’. Finally, the output after applying POS tagging is:

Wife /NN say/VBP canon/NN cameras/NNS are/VBP easy/JJ use/VB

Where NN and NNS present a noun and noun phrase, VBP encodes a verb (present tense), VB reflects a verb, and JJ represents an adjective.

4.1.2 Feature and Opinion Extractor

Feature and opinion extractor is responsible for the extraction of candidate features and opinion words. In the Opinion Analyzer, feature extraction depends on the nouns or noun phrases to produce a set of candidate features as actual product features are likely to be commented on by many customers.

In reviews people are more likely to discuss relevant features, which suggest that features must be frequent nouns. As discussed in Chapter two, existing research shows that 60-70% of the features are explicit nouns (Liu, 2006). The Opinion Analyzer makes the same assumption for the extraction of candidate features based on the hypothesis that product features are usually frequent nouns or noun phrases (Eirinaki et al., 2012; Miao et al., 2009). To extract frequent features, POS tags are utilized.

After extraction of frequent nouns, an opinion score (Oscore) is computed for every frequent noun. First, the opinion scores of the nouns are initialized to zero. For each associated opinion word, the *Oscore* of the noun is increased by one followed by the calculation of a title score (Tscore), which reflects the number of times the feature appeared in the reviews' titles. The feature frequency (FF), *Oscore* and *Tscore* are then summed up into a final score (Fscore). This final score (Fscore) is used to filter the nouns and identify potential features by selecting the nouns which have a score above a particular threshold. This threshold can be chosen based on experiments and human evaluation on different review data sets. After this, a file which contains the most frequent identified noun is built. The process is described in the following algorithm (NounScore).

Inputs: List of frquent nouns, Reviews

Output: Nouns with corresponding scores

NounScore (List of nouns, Reviews)

FF, Oscore, Tscore, Fscore 🡐0

  For every noun **do**

    FF🡐 frequency of noun in review document

    For every associated opinion word **do**

    Oscore 🡐Oscore+1

    End For

    For every review title **do**

If (review title contains the noun) **then**

Tscore $\leftarrow$ Tscore+1

End If

End For

Fscore $\leftarrow$ FF+ Oscore+ Tscore

End For

To extract opinion words, we used a window of size K, which means opinion words that are within K words of the selected feature are considered as associated opinion words, based on the assumption that an opinion word associated with a feature will be mentioned in its vicinity (Yang et al., 2010). After extraction of opinion words, a file called 'feature-opinion list' is maintained, which contains the identified features and corresponding opinion words with their opinion orientation and strength.

Consider the following review shown in Figure 4.11 which discusses three features: '*picture quality*', '*battery*' and 'viewfinder' with three opinion words: '*good*', '*poor*' and '*very good*' (presented in bold format) having opinion strengths +1, -1 and +2, respectively.

HR=80 **Rating** = ★★★★★

Title: Poor Battery

The **picture quality** is **good**
The **battery** is **poor**
It has **very good viewfinder**

Figure 4.11: An Example of Feature-Opinion List

The resulting feature-opinion list of the review shown in Figure 4.11 after applying the algorithm is presented below in Figure 4.12:

picture quality: good (+1)

battery: poor (-1)

viewfinder: very good (+2)

Figure 4.12: Feature-Opinion List

The next step after extraction of features and opinion words is to rank reviews using the review ranker as shown in Figure 4.7.

4.1.3 Review Ranker

For a given collection of reviews, review ranker ranks the reviews according to semantic and statistic clues, helpfulness ratio and users' rating according to the users' preferences. The inputs of the review ranker are a processed review document produced by the data pre-processor, feature-opinion list generated by the feature and opinion extractor and users' preferences (W1, W2,W3, W4, W5). The output of the review ranker is the ranking of reviews according to the users' preferences. The core of the review ranker is a RevRank algorithm which assigns weights to reviews based on four parameters (i) feature frequency, (ii) opinion words frequency, (iii) helpfulness votes, and (iv) users' rating. The proposed RevRank algorithm is stated below:

Inputs: Reviews, Users' Preferences (W1, W2,W3, W4,W5)

Output: Ranked Reviews

RevRank Algorithm (Reviews, W1, W2,W3, W4,W5)

```
For each Review do
    OWB, OWT, FB, FT, Rweight, Tscore, Maxweight, Rclass ← 0
        For review title do
            If (review title contains opinion words) then
            OWT ← number of opinion words in the review title
            End If
            If (review title contains features) then
            FT ← number of feature in the review title
            End If
        Tscore ← OWT + FT
        End For
        For each sentence in the body of the review do
            OWB ← OWB + number of opinion words in the line
            FB ← FB + number of features in the line
        End For
    Rweight ← ((W1* FB) + (W2* OWB) + (W3 * HR)+ (W4 * R))+ (W5*Tscore *α)
        For i ← 1 to number of review do
          If (Rweight> Maxweight) then
          Maxweight ← Rweight
          End If
        End For
    Rclass ← Rweight/ Maxweight*5
End For
```

In the proposed review ranking, two scores are calculated; one for the body of the review and the other for the title. For each review, the algorithm computes two scores for features: features are counted in the title (FT) and body (FR), separately. Similarly, opinion words are counted separately from the title (OWT) and body (OWB). Then, a score for the title of a review (Tscore) is computed using the number of features (FT) and the number of opinion words (OWT) in the title of a review. Finally, the review weight is calculated using the line 16 of the algorithm, where R is the users' rating, HR is the helpfulness ratio of the review. In order to incorporate users' preferences, the contribution of each parameter is defined by the values of W1, W2, W3, W4 and W5, which is provided by users. In line 16 of the algorithm, α is the title weight coefficient. Mostly, the title of a review is a good summary that captures the overall attitude of the reviewer towards the target product and thus should be given a larger weight, which is expressed by the title weight coefficient α. However, α is a parameter of the proposed algorithm and can be tuned appropriately depending on the data set on which the algorithm is applied. In this work, the value of α is set to 10 based on the experimental results of Eirinaki et al. (2012).

The review document may contain useless reviews having no significant impact on feature ranking (Chen & Tseng, 2010), therefore, detecting and filtering low-quality reviews may improve feature ranking (Liu et al., 2007). Consequently, inspired from the review quality categories of Chen & Tseng (2010) and Liu et al. (2007), the current study classified the reviews into five review quality classes: excellent, good, average, fair and poor. Informative reviews contain opinions on the features, hence they are often found to be more helpful. They also have a higher rating. The high quality reviews (i.e. excellent and good reviews) present in-depth opinions on product features in order to make them productive for opinion summarization. The medium-quality reviews (i.e. average and fair reviews) provide few opinions on products or features. The low-quality reviews may contain little information about a product/feature, or the information is too objective. It has been highlighted that the top-5 features account for 80% to 100% of high-quality reviews, therefore low-quality reviews can be excluded from opinion summary (Chen & Tseng, 2010; Liu et al., 2007). The filtration of low-quality reviews from the review document depends on the users' preferences that defines review class (es) utilized by the Opinion Analyzer for feature ranking, for instance, if a user selects excellent and good reviews then the feature ranking is applied only on these particular review classes.

As an example consider the following review shown in Figure 4.13 in which features and opinion words are highlighted in red and green colors, respectively. The *Tscore* of the review is two (1+1) as the title of the review contains one opinion word and one feature. The review discussed four features in the body (picture quality, view finder, battery and zoom), therefore the *OWB* score of the review is four. The *FB* score of the review is four as there are four opinion words in the review. Putting the values of these scores in line 16 of the algorithm results in 25.4 weight of the review shown in Figure 4.13, assuming the equal value of 0.20 for all users' preferences (W1, W2, W3, W4, and W5).

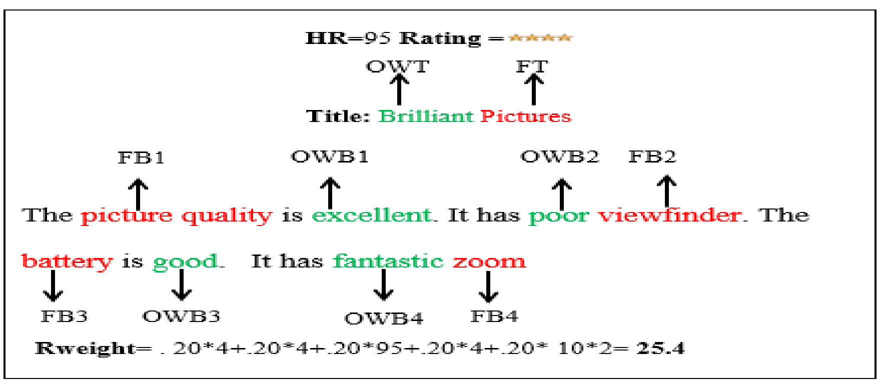


Figure 4.13: An Example of Review Weight Calculation

An example of review ranking process with three reviews is shown in Figure 4.14 in which a digital camera is regarded as an object.

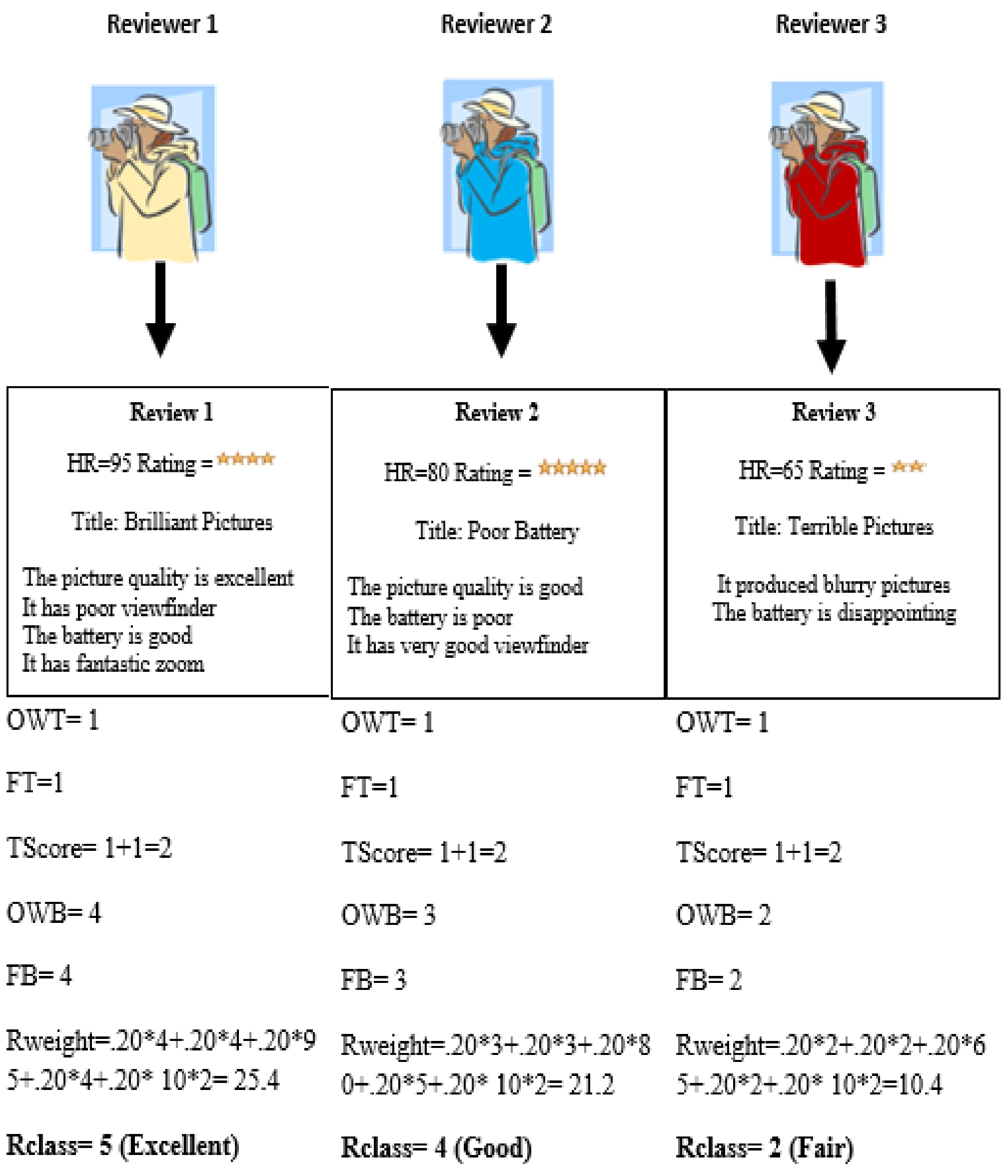


Figure 4.14: An Example of Review Ranking Process

4.1.4 Feature Ranker

After discarding low-quality reviews, the next step is to rank the extracted features by utilizing high-quality reviews. Feature ranker is responsible for feature ranking with the objective to discover all feature tuples (proposed in Section 4.1) from a collection of reviews and then to rank the extracted features. Inputs of the feature ranker are high-quality reviews identified by the review ranker and feature-opinion list produced by the feature and opinion extractor. The feature ranker generates ranking of features according to their corresponding semantic orientation and opinion strength as output.

First, a feature weight (FW) is assigned to the extracted features. Equation four is proposed in this work to calculated features weights where *FF* is the feature frequency, *OW* is the number of opinion words associated with the feature, *Rcount* is the number of reviews in which feature was expressed, and *Tcount* is the number of reviews' titles in which the feature was appeared. The *FW* is based on the idea that frequently discussed features in numerous reviews with more associated opinion words that appeared in more reviews' titles are decisive product features. *Tcount* and *Rcount* is utilized in computing the weight of a feature for the reason that if a feature is discussed in multiple reviews and reviews' titles, then it is a signifant feature.

$$FW = FF + OW + Rcount + Tcount \tag{4}$$

Secondly, the feature ranker calculates two ranks for every feature that is positive and negative ranks using opinion intensity. Every opinion word that appears in the selected reviews is assigned an empirical value manually (i.e. opinion intensity) that determines how positive or negative an opinion word is based on the classification provided in (Binali et al., 2009; Osimo & Mureddu, 2012). Basically, there are six classes: weakly positive (WP), mildly positive (MP), strongly positive (SP), weakly negative (WN), mildly negative (MN) and strongly negative (SN). The opinion intensity ranges between +1(WP) to -3(SN). For instance, +3 is assigned to '*excellent*' and +1 is assigned to '*good*'. Similarly, -3 is assigned to '*terrible*' and -1 is assigned to '*bad*'.

Positive rank (P_{rank}) of a feature reflects the number of positive opinion words associated with a feature and the accumulated strength of associated positive opinion words. Equation 5 is proposed to calculate P_{rank} of a feature that is shown below:

$$P_{rank} = POW + \sum_{i}^{m} OWP_i \sum_{j}^{m} OMP_j = \sum_{k}^{o} OSP_k \tag{5}$$

In equation five, OWPi is the occurrences of every weakly positive opinion word, OMPj is the occurrences of every mildly positive opinion word and OSP_k is the occurrences of every strongly positive opinion word. The larger the value of P_{rank}, the more positively the selected feature was discussed by the users.

The negative rank (N_{rank}) of a feature encodes the number of negative opinion words associated with a feature and the accumulated strength of associated negative opinion words. Equation 6 is proposed to calculate the N_{rank} of a feature that is shown below:

$$N_{rank} = NOW + \sum_{i}^{m} OWN_i \sum_{j}^{m} OMN_j = \sum_{k}^{o} OSN_k \tag{6}$$

where OWN_i, OMN_j and OSN_k are the occurrences of every weakly negative, mildly negative and strongly negative opinion words, respectively. The intuition behind the N_{rank} is that if a feature is described by more negative words then it should be ranked higher than others.The larger the value of N_{rank}, the more negatively the feature had been discussed by the users.

The proposed feature ranking process is illustrated in Figure 4.15. A review document (R_s) contains a set of reviews. Every review (R_j) belongs to R_s contains a set of feature-opinion pair (F_{ij}, O_{ij}), which comprises of a feature F_{ij} and an opinion word O_{ij}. The pair (F_{ij}, O_{ij}) is employed for the proposed feature ranking. I_{ij} represents the importance (weight) of feature F_i in review R_j. The P_{rank} of F_{ij} is derived by a feature ranking process from positive opinion words (P_{OWij}) and positive opinion strength (WP, MP, SP, WN,MN, SN) where WP is the weakly positive opinion with opinion strength +1, MP is the mildly positive opinion with opinion strength +2 and SP is the strongly positive opinion with opinion strength +3. Similarly, negative opinion words (N_{OWij}) and negative opinion strength (WN is the weakly negative opinion with opinion strength -1, MN is the mildly negative opinion with opinion strength -2 and SN is the strongly negative opinion with opinion strength -3) are utilized to calculated N_{rank} of feature $F_{ij.}$ Then the P_{rank} and N_{rank} of a feature are summed into an overall rank (O_{rank}) to derive a whole rank for the feature as shown in equation 7.

$$O_{\mathbf{rank}} = P_{\mathbf{rank}} - N_{\mathbf{rank}} \tag{7}$$

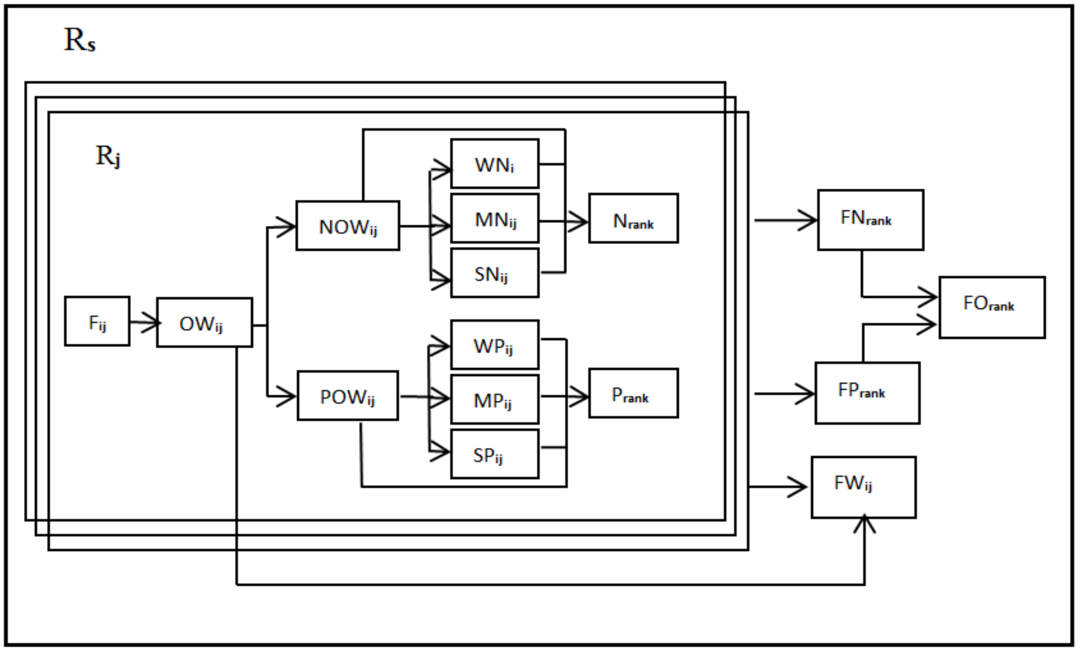


Figure 4.15: Proposed Feature Ranking Process

The mathematical model of the proposed method is presented below:

$Rs = \{R1, R2, R3, \dots\dots\dots., Rn\}$

$\forall\, Rj \in Rs,$

$Rj = \{(f1j, o11j), (f1j, o21j), \dots, (f1j, om1j), (f2j, o12j), \dots, (f2j, o22j), \dots, (f2j, om2j),$

$(fnj, o1nj), \dots, (fnj, o2nj), \dots, (fnj, omnj)\}$

$f = \{f1, f2, f3, \dots\dots., fn\}$

$OW = \{NOW, POW\}$

$POW = \{WP1, WP2, \dots, WPn, MP1, MP2, \dots, MPn, SP1, SP2, \dots, SPn\}$

$NOW = \{WN1, WN2, \dots, WNn, MN1, MN2, \dots, MNn, SN1, SN2, \dots, SNn\}$

The feature ranker is based on the proposed FRank algorithm, which calculates the FW, P_{rank}, N_{rank} and O_{rank} of prominents features. The input of the algorithm is the feature-opinion list and the output of the algorithm is the ranking of features. The proposed FRank algorithm is stated below:

Input: List of Feature, Reviews

Output: Ranked Features

FRank algorithm (Features List)

```
For each feature do
    FW, SP, MP, WP, SN, MN, WN, FTScore, OWScore , RCount ← 0
    POW, NOW, Prank, Nrank, Orank ← 0
        For each review do
            If (the feature appeared in the review title) then
                TScore ← TScore +1
            End If
            If (the feature appeared in the review) then
                RCount ← RCount +1
            End If
        For each strongly positive opinion word associated with the feature do
            SN ← SN + 1
            POW ← POW + 1
        End For
        For each mildly positive opinion word associated with the feature do
            MN ← MN + 1
            POW ← POW + 1
        End For
        For each weakly positive opinion word associated with the feature do
            WN ← WN + 1
            POW← POW+ 1
        End For
        For every strongly negative opinion word associated with the feature do
            SN ← SN + 1
            NOW ← NOW + 1
        End For
        For each mildly negative opinion word associated with the feature do
            MN ← MN + 1
            NOW ← NOW + 1
        End For
        For each weakly negative opinion word associated with the feature do
            WN ← WN + 1
            NOW ← NOW+ 1
        End For
End For
For each associated opinion word do
    OWScore ← OWScore + 1
End For
FF= Feature Frequency
FW= FF + OWScore + TScore+ RCount
Prank ← POW+3*SP+2*MP+WP
Nrank ← NOW+3SN+2*MN+WN
Orank ← Prank – Nrank
```

The feature list is composed of prominent features of a target product. The feature weight (FW), strongly positive opinion words (SP), mildly positive opinion words (MP), weakly positive opinion words (WP), strongly negative opinion words (SN), mildly negative opinion words (MN), weakly negative opinion words (WN), title score of the feature (FTScore), opinion word score of the feature (OWScore), associated positive opinion words (POW), associated negative opinion words (NOW), P_{rank}, N_{rank} and O_{rank} are calculated for each feature. *TScore* represents the number of titles in which the feature appeared and *RCoun*t represents the number of reviews in which the feature appeared. The initialization of variables for each feature in the feature list is followed by counting each opinion word in related variables (OWScore, POW, NOW, SP, MP, WP, SN, MN, WN) and the frequency of features in related variables (FF, TScore, RCount). For each review, separate scores for strongly positive opinion word, mildly positive opinion word, weakly positive opinion word, strongly negative opinion word, mildly negative opinion word and weakly negative opinion word associated with each of the features is determined. Finally, P_{rank}, N_{rank} and O_{rank} are calculated for each review.

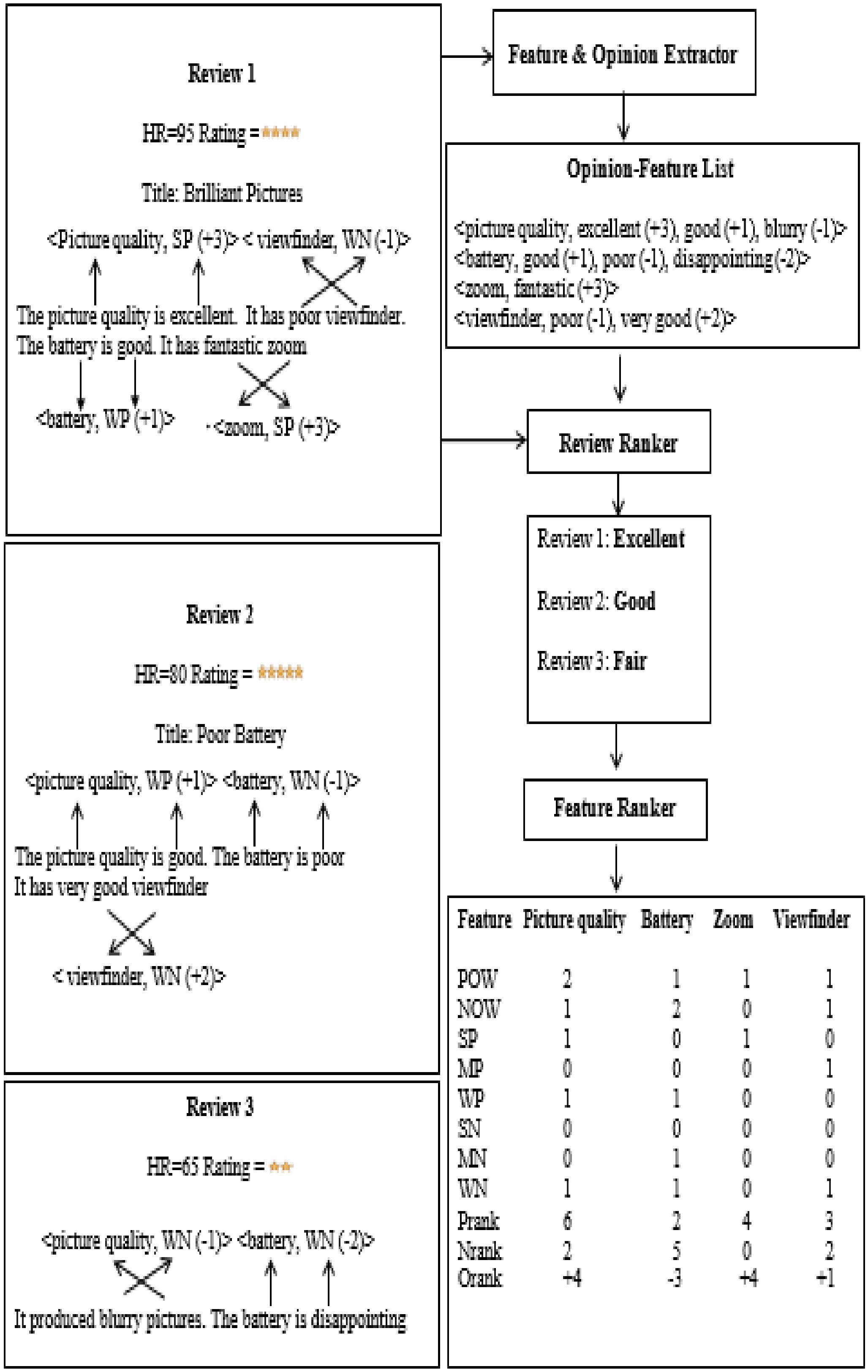


Figure 4.16: An Example of Proposed Feature Ranking

In Figure 4.16, the reviews are expressing opinions on features ‘*picture quality*’, ‘*battery*’, ‘*zoom’* and ‘*viewfinder*’ of a camera. In these reviews, the feature ‘*picture quality*’ is described by two positive opinion words (excellent, good) and one negative opinion word (blurry). Therefore, the POW and NOW counts of the feature ‘*picture quality*’ are two and one, respectively. The opinion words ‘*excellent*’ ,‘*good*’ and ‘*blurry’* are strongly positive (+3), weakly positive (+1) and mildly negative (-2), respectively, hence the SP, WP, MN counts are set to one. One positive opinion word ‘*good*’ and two negative opinion words ‘*poor*’ and ‘*disappointing*’ are associated with the feature ‘*battery*’, thus, the resulting values of the POW and NOW are one and two, respectively. The opinion strength of the opinion words are weakly positive (+1), weakly negative (-1) and mildly negative (-2), respectively, consequently, the WP, WN and MN counters are set to one. The feature *‘zoom’* is defined by the opinion word *‘fantastic’,* which is a strongly positive opinion word with an opinion strength of +3, as a result the value (one) is assign to the POW and SP counters of the feature. One weakly negative opinion word *‘poor’* and one mildly positive opinion word ‘*very good*’ are connected with the feature ‘*viewfinder*’, subsequently the POW, NOW, WN are MP set to one. All other counters are set to zero. Finally, the values of POW, NOW, SP, MP, WP, SN, MN, WN are utilized to compute P_{rank} and N_{rank} of each feature and then P_{rank} and N_{rank} are summed up to calculate O_{rank} of a feature using equation five, six and seven as shown in Figure 4.16. The P_{rank} , N_{rank} and O_{rank} of the feature ‘*picture quality*’ are calculated as shown in Table 4.1, 4.2 and 4.3.

Table 4.1: P_{rank} Calculation

Feature	POW	WP	MP	SP	P_{rank}
Picture Quality	2	1	0	1	2+1*1+0*2+1*3=6

Table 4.2: N_{rank} Calculations

Feature	NOW	WN	MN	SN	N_{rank}
Picture Quality	1	1	0	0	1+1*1+0*2+0*3=2

Table 4.3: O_{rank} Calculations

Feature	P_{rank}	N_{rank}	O_{rank}
Picture Quality	6	2	6-2=4

4.1.5 Opinion Visualizer

In order to identify the best opinion visualization technique for the Opinion Analyzer, a questionnaire survey was conducted to get the users' preferences about existing opinion visualization techniques. Based on the results of the survey, users' preferred techniques will be used for displaying feature ranking and an opinion-strength-based summary. The following sub-sections (from Section 4.1.51 to Section 4.1.5.3) describes the conceptual design and the research methodology used in the survey.

4.1.5.1 Instrument

A questionnaire survey consisting of ten structured close-ended questions was developed to collect the data. The ten evaluation metrics (and the assessment areas) were adapted from (Bai, White, & Sundaram, 2011) (see Table 4.4). Six out of seven assessment areas are selected because it was difficult to collect data for the seventh assessment area, which is about domain experts' opinion. Each assessment area defined a set of evaluation categories and sub-categories (metrics). The questionnaire had two parts. Part A required the participants to fill in their demographic profiles, such as, age and gender. Part B required the participants to state their level of agreement or disagreement about visualizations against each metric on a Likert scale that ranges from Strongly Disagree (1) to Strongly Agree (5) (see Table 4.4).

Table 4.4: Questionnaire Part B with Metrics and Assessment Areas

Assessment Area	Metrics	Questions	1	2	3	4	5
Visual Impact	Visual Appeal	Q1: The visualization is visually appealing					
Overall Performance	Easy To Understand	Q2: The visualization is easy to understand.					
	User Friendly	Q3: The visualization is user friendly.					
Overall Design	Informativeness	Q4: The visualization is informative.					
	Intuitiveness	Q5: The visualization is intuitive.					
Information Quality	Usefulness	Q6: The visualization is useful.					
	Comprehensiveness	Q7: The comprehensiveness of data is good.					
Visual Representation	Comparison Ability	Q8: The comparison of data is good.					
	Representation Style	Q9: The representation style of data is good.					
Information Presentation Model	Pre-Knowledge Required	Q10:Pre-Knowledge is required to understand the visualization.					

The questionnaire was utilized to collect data about existing feature-based opinion visualization techniques, including, opinion wheel (Wu et al., 2010), rose plot (Gregory et al., 2006), positioning map (Morinaga et al., 2002), line graph and pie chart (Miao et al., 2009), comparative relation map (Xu et al., 2011), tree map (Gamon et al., 2005), visual summary (Oelke et al., 2009), glowing bar (Gamon et al., 2008), bar chart with symbols (Wanner et al., 2009) and bar chart (Liu et al., 2005) based on intuitiveness, complexity and level of abstraction. These techniques are already discussed in Section 3.4.

4.1.5.2 Instrument Refinement

The questionnaire was pre-tested by conducting a pilot study to judge its feasibility. The pilot study was performed with the help of a focus group to gain the participants' understanding about the questionnaire. There were fifteen participants in the focus group study, five in each group. The participants were academics from COMSATS Institute of Science and Technology, Islamabad, Pakistan. The participants discussed the questionnaire with each other and provided their level of understanding, suggestions and comments. Initially 14 metrics were selected, however, after the pilot study four metrics, namely, eye pleasing, visually uncomfortable, stunning, and conciseness were removed as the first three metrics can be represented by visual appeal. Similarly, conciseness is closely related to the comprehension metric. Then, the refined questionnaire was again discussed with two participants of the focus group study to finalize its contents and the finalized questionnaire was used to collect data.

4.1.5.3 Data Collection

Seminars and a web-based online questionnaire were used to collect the data.

- **Seminar**

The author of the thesis contacted the coordinator of Computer Science Department of COMSAT Institute of Information Technology, Pakistan, for conducting seminars on opinion visualization techniques. The coordinator made arrangements for the seminars and informed the author, students and academics about the venue and the time of the seminars. The target participants of these seminars are students and academics with a computer science background, especially students who took the 'Human Computer Interaction' course. This course introduced some of the prominent visualization techniques, such as tree map. As a result, the students who took the course had a better knowledge about visualization techniques. The reason behind targeting computer science personnel is, they are the largest group of Internet users and are more likely to consult and use online opinion information than other Internet user groups.

Three seminars were conducted to present the information about state-of-the-art opinion visualization techniques and for a dynamic discussion about the visualization techniques with target participants.The author of the thesis gave a ten-minute presentation at the beginning of each seminar in order to present the objectives of the study, a brief introduction of selected opinion visualization techniques and instructions on filling the questionnaire, so that the participants had an adequate understanding to fill in the questionnaire. Then, a question-answer session was held in order to clarify the understanding of the participants about the visualization techniques. In the session, the participants asked mostly the questions about the meaning of a symbol or metaphor in the visualizations such as the interpretation of symbols in Figure 3.9 (Section 3.4.4). Approximately, the session lasted for 30 minutes. After the session, the opinion visualization techniques were displayed and the participants were requested to fill in the questionnaire. After that an interactive discussion was held between the participants and the presenter in which the participants provided their preferences about opinion visualizations. The preferences were noted down. Approximately, 60 minutes were consumed in the first seminar. The same procedure was repeated for second and third seminars.

- **Online Questionnaire**

To increase the number of responses, an online questionnaire on https://www.limeservice.com/en/ was also created and the link was distributed via e-mail to computer science students at different universities in Malaysia. A video was added in the online questionnaire that briefly introduced the selected opinion visualizations. One of the limitations of the online questionnaire is the lack of face-to-face communication, especially a question answer session to clarify concepts. To overcome this limitation, a description of concepts, metaphors and symbols was added in the online questionnaire, which was asked by the participants of the seminars in the question-answer session. It took approximately five weeks to collect the data.

4.1.5.4 Participants

A total of 146 users participated in the data collection. The participant's size of the seminar was 110 (22 females and 88 males). The participant's size for online questionnaire was 36 (17 females and 19 males). Table 4.5 shows the details of the participants (M: 25.57, SD: 5.55). Most of the participants belonged to the 21-30 years age group as they used more

Internet than other categories. A large number of participants have prior experience of getting decision oriented information from online reviews.

Table 4.5: Details of the Participants

	Categories	Number
Gender	Male	116
	Female	30
Age	<20	18
	21-30	113
	31-40	11
	>40	4

4.2. Evaluation of Opinion Analyzer

The Opinion Analyzer was implemented in Python 2.7 using natural language toolkit (NLTK). A real data set from amazon.com was used for the evaluation of the proposed review and feature ranking methods, and a usability investigation was carried out to access the usefulness of opinion-strength-based visualization. The evaluation of the Opinion Analzer was conducted by measuring the accuracy of review quality evaluation, P_{rank}, N_{rank} and O_{rank}.

Table 4.6 shows an example how the accuracy of proposed review ranking method is calcualted for two reviews. The class of each review is calculated mannually (actual class) and then compared with system generated (extracted) class to compute the accuracy of reviews using the formula given in Section 2.8.1. The accuracy reflects that how accurate the proposed review ranking method is in assigning the review quality class.

Table 4.6: Accuracy Calculation of Proposed Review Ranking Method

Reviews	Actual Class	Extracted Class	Accuracy
Review 1	Good (4)	Average (3)	3/4*100=75 %
Review 2	Average (4)	Average (4)	4/4*100=100 %

Similarly, the manually calculated values of P_{rank}, N_{rank} and O_{rank} were compared with extracted (system generated) values to obtain the accuracy of the proposed feature ranking method using the formula discussed in Section 2.8.1. Table 4.7 shows an example how the accuracy of P_{rank}, N_{rank} and O_{rank} were computed.

Table 4.7: Accuracy Calculation of P_{rank}, N_{rank} and O_{rank}

	Actual P_{rank}	Extracted P_{rank}	Accuracy of P_{rank}	Actual N_{rank}	Extracted N_{rank}	Accuracy Of N_{rank}	Actual O_{rank}	Extracted O_{rank}	Accuracy of O_{rank}
Picture Quality	15	12	12/15*100= 80%	10	9	9/10*100 =90%	5	3	3/5*100 = 60%

The following sections describe the details of experimental data set and setup along with the usability study conducted in this research work.

4.2.1 Experimental Data Set and Setup

For the evaluation of the Opinion Analyzer, experiments were conducted on a real data set from amazon.com taken from Prof. Bing Liu. This data set is used by multiple studies in the past including Hu & Liu (2004), Ding et al. 2008, Qui et al. (2009), and Jian et al. (2010). Hence this data set is considered to be a benchmark in opinion mining. The data set consists of pre-processed customer reviews of five products: two digital cameras (Canon PowerShot G3 and Nikon Coolpix 4300), one cellular phone (Nokia 6610), one MP3 player (Creative Labs Nomad Jukebox Zen Xtra 40GB) and one DVD player (Apex AD2600 Progressive-scan DVD player). Table 4.8 shows the details of the data set.

Table 4.8: Detailed Information of Data Set

	Product Name	**Number of Reviews**	**Number of Sentences**	**Length in Words**	**Length in Characters**
1.	Digital camera 1	45	597	11280	48714
2.	Digital camera 2	34	346	6749	29763
3.	Cellular phone	44	546	9681	42795
4.	MP3 player	95	1716	12719	54872
5.	DVD player	100	740	32553	138301
	Total	318	489	72982	314445

The Opinion Analyzer was implemented in Python 2.7 using natural language toolkit (NLTK). The evaluation of the Opinion Analzer was conducted by measuring the accuracy of the review quality evaluation, P_{rank}, N_{rank} and O_{rank} for the five products. The Opinion Anaylzer is deterministic in nature, therefore the same results were produced every time the Opinion Analyzer was executed. However, the Opinion Analyzer was run five times for each data file to check the consistency of results.

4.2.2 Usability Study

In order to access the usability of the proposed opinion-strength-based visualization, a user study with ten participants (4 Female, 6 Male) was conducted. The participants are postgraduate students at University of Malaya. The purpose of the study is to verify the effectiveness of the visualization. First, the participants were given a five minute introduction of the underlying concepts of the visualizations. Then, the participants were asked to state their level of agreement or disagreement about visual appeal, understandability, user-friendliness, informativeness, and intuitiveness of the visualization on a Likert scale that ranges from Strongly Disagree (1) to Strongly Agree (5).

Chapter 5 : Results and Discussion

This chapter presents the results of the questionnaire survey to highlight the users' preferences about existing opinion visualization techniques. The results of the experiments carried out by the Opinion Analyzer on a real data set of five electronic products from amazon.com (discussed in Section 4.2.1) are also described in this chapter along with opinion-strength-based visualizations for each product. Moreover, the findings of the usability study are presented in this chapter.

5.1 Experimental Results and Discussion

This section presents the results of review quality evaluation, P_{rank}, $N_{rank,}$ O_{rank} and opinion-strength-based visualization for each product.

5.1.1 Review Quality Evaluation

To evaluate the effectiveness of proposed review ranking method (Reseach Objective 4), we utilized the matric accuracy as mentioned in section 2.8.1. The review weight of each review is calculated using the equation presented in line 16 of the RevRank algorithm stated in Section 4.1.3. Then the review quality class of each review is calculated according to their corresponding weights.

This section presents the reviews quality classfication of all the data files along with the accuracy achieved by the Opinion Analyzer.

5.1.1.1 Review Quality Evaluation for Digital Camera 1

Figure 5.1 presents the quality of reviews of the product 'Digital Camera 1'. X-axis represents the number of reviews in each class and Y-axix represents the review quality classes: excellent, good, average, fair and poor as discussed in Section 4.1.3.

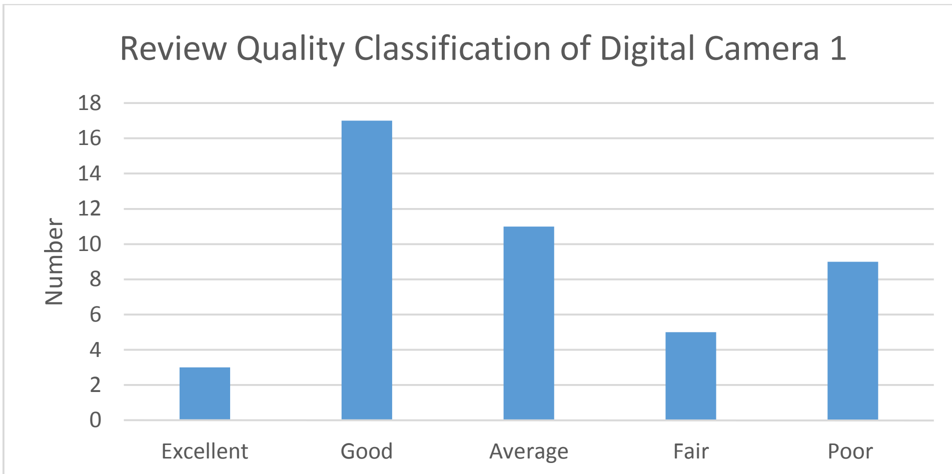


Figure 5.1: Reviews Quality Classification of Digital Camera 1

It can be inferred from Figure 5.1 that the majority of the product reviews are 'Good' providing sufficient amount of information about features and customers' opinions. Only 31% reviews were categorized as 'fair' and 'poor'. It is notable that only 7% reviews belong to 'Excellent' review class so in this case approximately 69% of reviews belong to 'Excellent, 'Good' or 'Average' class. Therefore, the classification of the reviews according to their qualities can be used to discard low quality reviews, and feature ranking can be applied only on the selected review quality class(es). For instance, excellent and good reviews can be selected for feature ranking and the other reviews might be discarded.

5.1.1.2 Review Quality Evaluation for Digital Camera 2

The reviews quality classification for Digital Camera 2 is depicted in Figure 5.2.

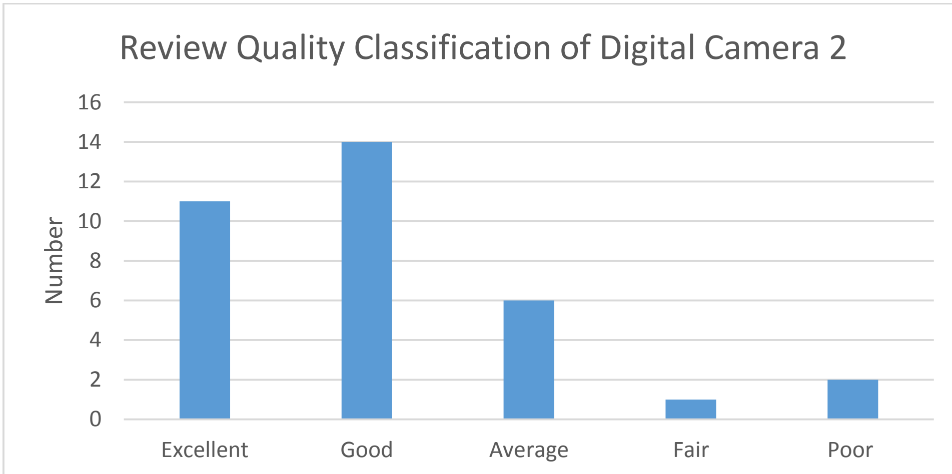


Figure 5.2: Reviews Quality Classification of Digital Camera 2

Similar to the 'Digital Camera 1', most of the reviews of 'Digital Camera 2' belong to 'Good' review quality class, however, it is interesting to note that a number of reviews also belong to 'Excellent' review class as well. Together, 'Excellent' and 'Good' reviews make 73% of the total reviews. Further, there are some reviews (17%) that belong to 'Average' review class. Only 8% reviews were categorized as 'fair' and 'poor'. Hence, the low quality reviews can be rejected after classifying the reviews according to their quality because 'Excellent', 'Good' and 'Average' review make 92% of the total reviews.

5.1.1.3 Review Quality Evaluation for Cellular Phone

Figure 5.3 exhibits the reviews quality classification for 'Cellular Phone'. Unlike 'Digital Camera 2', most of the reviews of 'Cellular Phone' belong to 'Poor' review quality class. It is important to note that in this case only 12% reviews exist in 'Exellent', 'Average' and

‘Fair’ classes with no review in ‘Good’ class. Therefore, it is deduced that a large number of the reviews are not providing enough decision-oriented information because most of the reviews (39 to be exact) fall in the "poor" category. Poor reviews imply that very few features are discussed in the review. They also lack effective opinion oriented information.

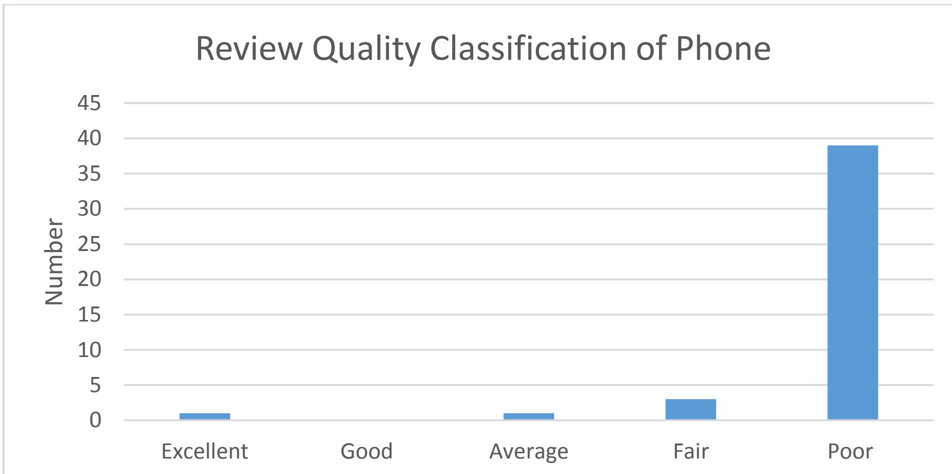


Figure 5.3: Reviews Quality Classification of Cellular Phone

5.1.1.4 Review Quality Evaluation for MP3 Player

The quality of reviews for ‘MP3 Player’ is shown in Figure 5.4 indicating that most of the reviews (42%) belong to ‘Good’ and ‘Execllent’ review quality classes collectively. Approximately 15% of reviews belong to ‘Average’ review quality class and only 5% reviews are categorized in ‘Fair’ class. However, there are 38% reviews that belong to ‘Poor’ review quality class. Therefore, it can be inferred that the majority of the product reviews (62%) are ‘Excellent’, ‘Good’ , ‘Average’ and ‘Fair’ providing sufficient amount of information about features and their corresponding opinions.

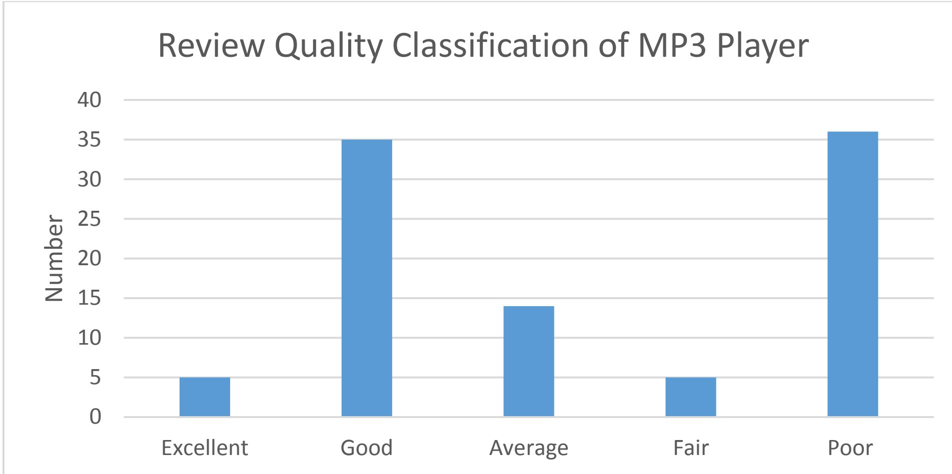


Figure 5.4: Reviews Quality Classification of MP3 Player

5.1.1.5 Review Quality Evaluation for DVD Player

Figure 5.5 describes the review quality classification for 'DVD Player'. Most of the reviews (44%) belong to the 'Excellent' and 'Good' review quality classes collectively. It can be deduced that the majority of the product reviews (58%) are either 'Excellent', 'Good' or 'Average' presenting adequate opinion information about different features of the product. However, there are 36% and 6% reviews that belong to the 'Poor' and 'Fair' quality classes, respectively.

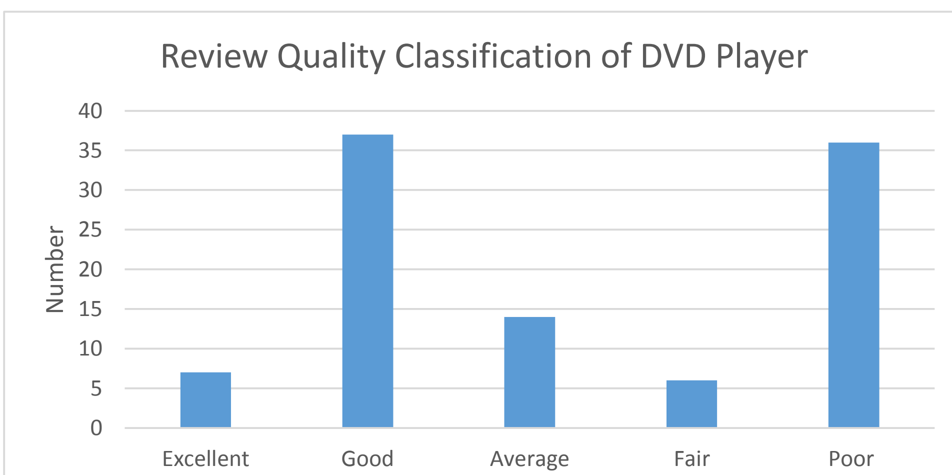


Figure 5.5: Reviews Quality Classification of DVD Player

Figure 5.6 illustrates the average accuracy of review quality classification for all the products. The Opinion Analyzer achieved 88% and 87% accuracy for the products 'Digital Camera 1' and 'Digital Camera 2' respectively, for the review quality classification. The product 'Phone' and 'DVD Player' are also up to mark with 85% and 83% accuracy, respectively. The Opinion Analyzer attained 81% accuracy for the product 'MP3 Player' for the review quality classification, hence, not lagging way behind the other products. The accuracy of the Opinion Analyzer for all the products was found to be 85%.

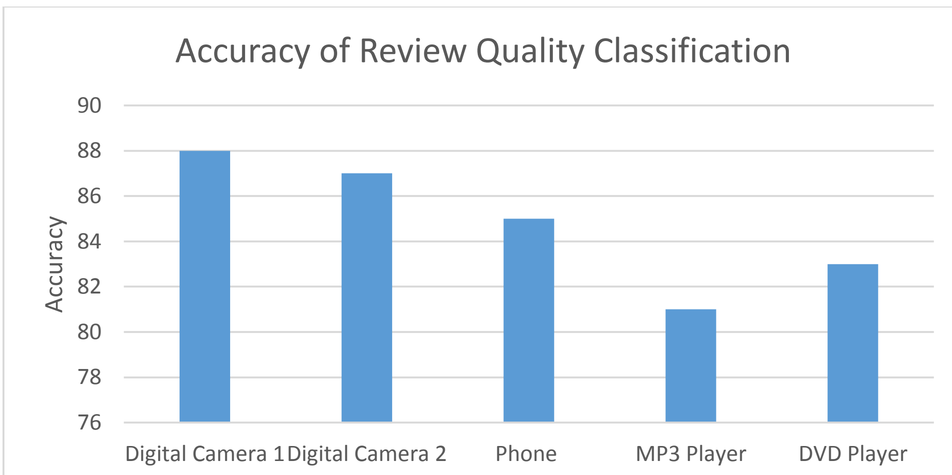


Figure 5.6: Accuracy of Reviews Quality Classification

5.1.2 Feature Ranking

To evaluate the effectiveness of the proposed feature ranking method (Reseach Objective 4), we utilized the metric accuracy as mentioned in section 2.8.1. This section presents the findings of P_{rank}, N_{rank} and O_{rank} of the five experimental data files and reports the accuracy achieved by the Opinion Analyzer for these ranking. Higher P_{rank} of a feature indicates that reviewers discussed these features more positively as compared to other features, however, higher N_{rank} reflects that reviewers dislike the feature and expressed more negative opinions on the feature. Overall opinion about a feature is expressed by its O_{rank} that integrates P_{rank}, and N_{rank}. Top ten features according to P_{rank} highlights the strengths of a product. On the other hand, N_{rank} indicates shortcomings of a product. The P_{rank}, N_{rank} and O_{rank} of each feature were calculated using equations 5, 6 and 7 discussed in Section 4.1.4. The manually calculated values of P_{rank}, N_{rank} and O_{rank} were compared with extracted (system generated) values to obtain the accuracy of the proposed feature ranking method using the formula discussed in Section 2.8.1.

5.1.2.1 Feature Ranking of Digital Camera 1

This section presents the top ten features of the product 'Digital Camera 1' according to the P_{rank}, N_{rank} and O_{rank}.

Figure 5.7 represents the top ten features of the product 'Digital Camera 1' according to the P_{rank} that highlights the strengths of the camera.

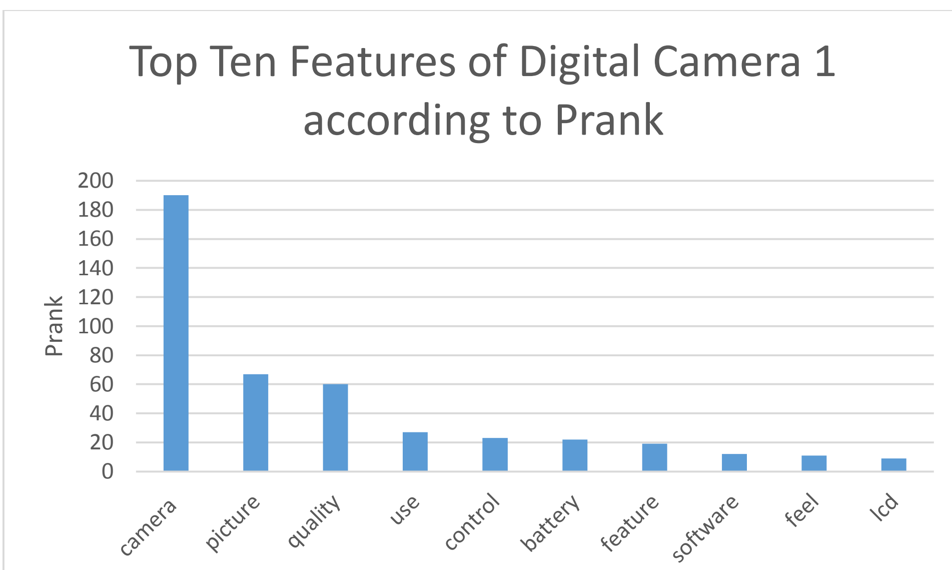


Figure 5.7: Top 10 Features of Digital Camera 1 according to P_{rank}

The overall feedback of the camera is remarkably positive as the P_{rank} of the camera is above 180. The feature 'Picture' is at the second position in the ranking with a P_{rank} of 67, representing that consumers appreciated the picture quality of the camera to a large extent. The P_{rank} of the feature 'quality' indicates that large number of users is satisfied with the quality of the camera. Higher P_{rank} reflects that more positive opinion words are associated with this feature. However, very few users are satisfied with the 'Software', 'Feel, and 'LCD' features of the camera as indicating the by P_{rank}.

Figure 5.8 shows the accuracy of the top 10 features of 'Digital Camera 1' according to the P_{rank} indicating that five of the top ten features resulted in 100% accuracy, namely, battery, feature, software, feel and LCD. According to Sun et al. (2013), if the accuracy is 100% for half data set then it can be considered as a good accuracy. Therefore, good accuracy is achieved by the Opinion Analyzer for the top ten feature of 'Digital Camera 1'.

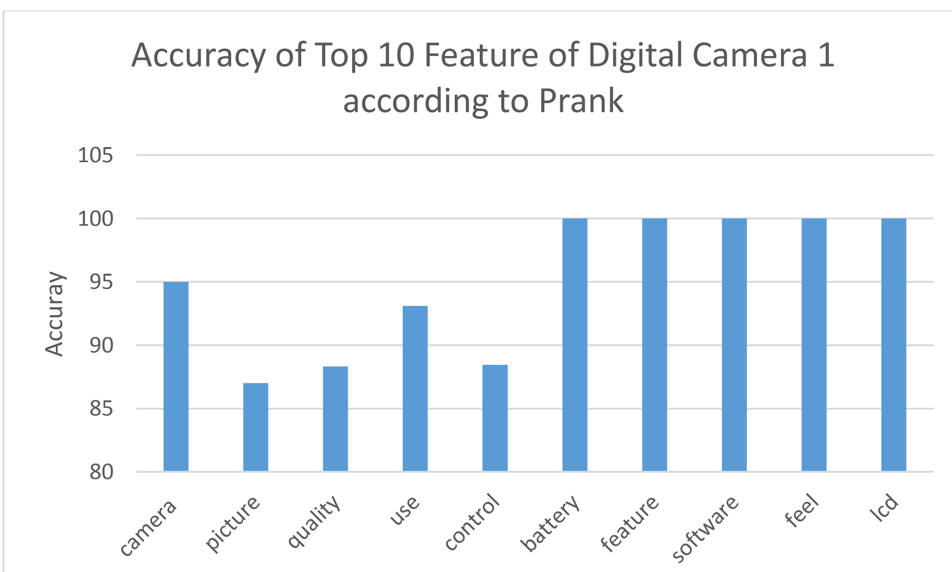


Figure 5.8: Accuracy of Top Ten Features according to P_{rank}

The features 'Camera' and 'Use' having 95% and 93% accuracy, respectively, however, the feature 'Picture' has the lowest accuracy of 87% resulted in the overall accuracy of 95% achieved by the Opinion Analyzer for P_{rank} of digital camera 1.

Figure 5.9 highlights the top 10 features of the digital camera 1 according to the N_{rank} that pinpoints the shortcomings and limitations of the camera.

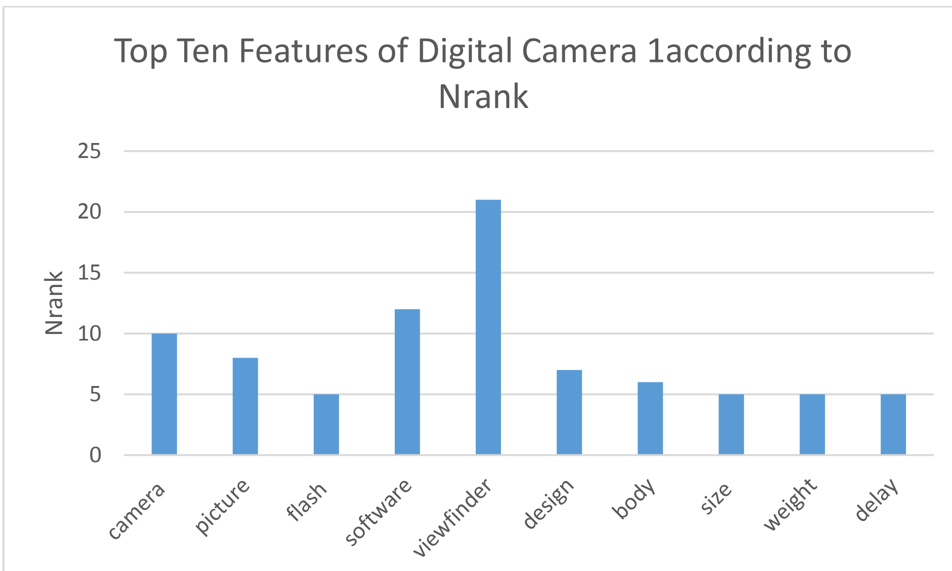


Figure 5.9: Top 10 Features of Digital Camera 1 according to N_{rank}

The feature 'Veiwfinder' has the highest N_{rank} score indicating that many users are not satisfied with the viewfinder of the camera, followed by 'Software, 'Camera' and 'Picture' features showing customers' dissatisfaction to a large extend. The features 'Design' and 'Flash' have N_{rank} greater than five representing that very few users are not satisfy with the design and flash of the camera. The features ‘Flash’, ‘Size’, ‘Weight’ and ’Delay’ have N_{rank} five indicating that very few users showed disappointment on these features.

Figure 5.10 shows the accuracy of the top 10 features of the digital camera 1 according to the N_{rank} indicating that good accuracy is achieved by the Opinion Analyzer as six of the top ten features received 100% accuracy, namely, picture, software, body, size, weight and delay. The next two features are ‘Viewfinder’ and ‘Camera’ having 91% and 83% accuracy, respectively. Greater than 60% accuracy was attained by the Opinion Analyzer for the features ‘Flash’ and ‘Design’ resulting in 91% overall accuracy of the Opinion Analyzer according for N_{rank}.

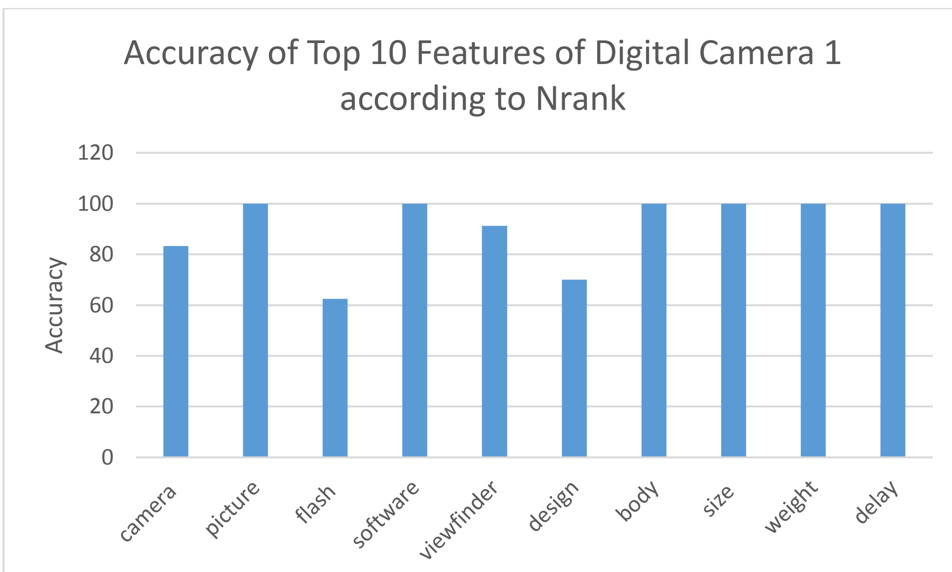


Figure 5.10: Accuracy of Top Ten Features according to N_{rank}

Figure 5.11 represents the top ten features of ‘Digital Camera 1’ according to the O_{rank} that combines P_{rank} and N_{rank} by highlighting the overall score (semantic) of the camera and features. Overall, the camera received O_{rank} score of 180 indicating the overall semantic about the camera is very positive. The feature 'Picture' and 'Quality' have O_{rank} score of 59 and 55, respectively, highlighting that many users commented on these feature positively, followed by 'Use', 'Control', 'Battery' and 'Feature' features showing that these features are also considered by users positively. The other important features on which customers showed their satisfaction are 'Feel', 'LCD' and 'Color'.

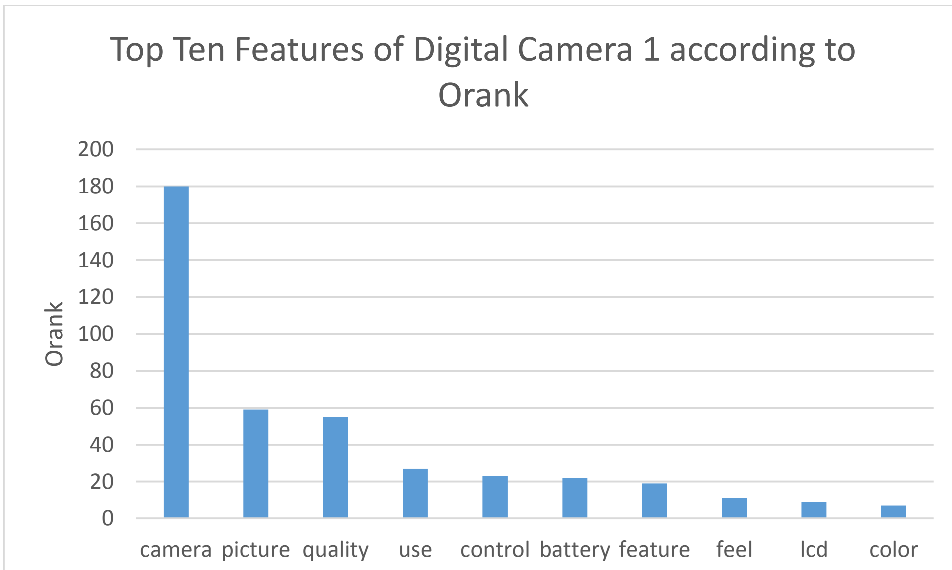


Figure 5.11: Top 10 Features of Digital Camera 1 according to O_{rank}

Figure 5.12 shows the accuracy of the top 10 features of the digital camera 1 according to the O_{rank} showing that the features 'Battery', 'Feature', 'Feel', 'LCD', and 'Color' received 100% accuracy.

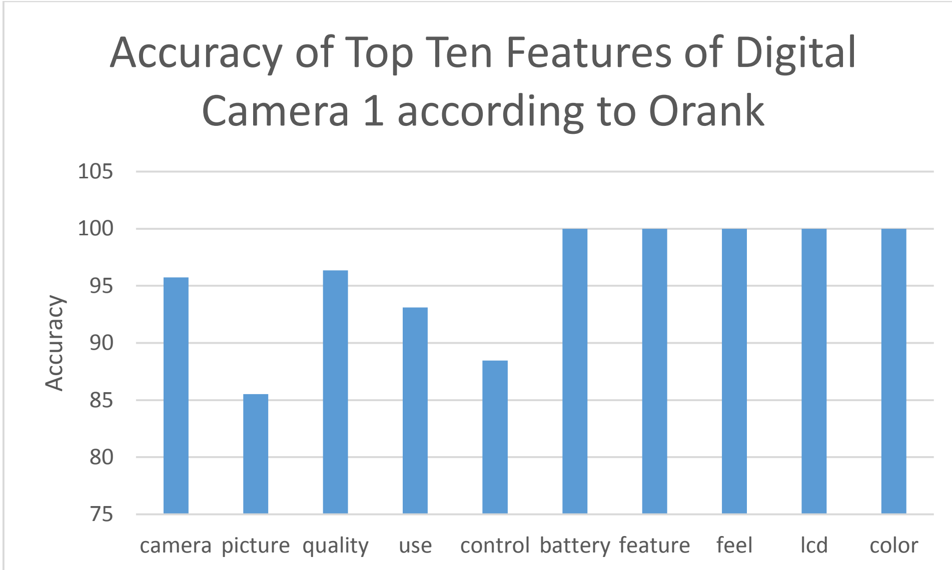


Figure 5.12: Accuracy of Top Ten Features of Digital Camera 1 according to O_{rank}

The next two features are 'Quality' and 'Camera' having 96% and 95% accuracy, respectively. The features 'Use', 'Control', and 'Picture' have accuracy greater than 85% resulting in 95% overall accuracy of the Opinion Analyzer according to O_{rank}.

5.1.2.2 Feature Ranking of Digital Camera 2

Figure 5.13 highlights the strengths of 'Digital Camera 2' by showing the top ten according to the P_{rank}. The P_{rank} of the camera is above 140 indicating that a large number of customers are satisfied with the camera.

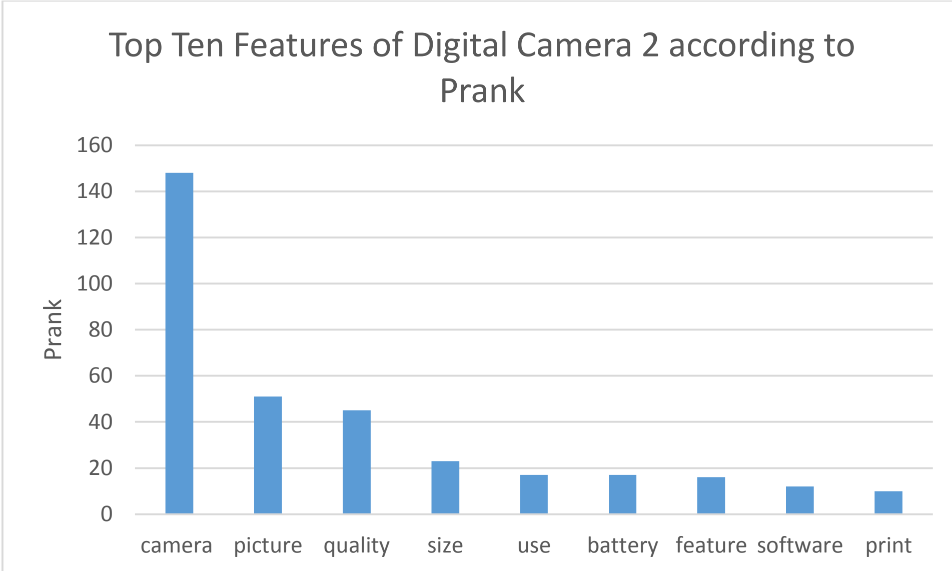


Figure 5.13: Top 10 Features of Digital Camera 2 according to P_{rank}

The other main features showing positive customers' opinions are 'Size', 'Use', 'Battery' and 'Feature' having P_{rank} in the range between 23 and 17. However, the features 'Software' and 'Print' are commented positively by very few users.

Figure 5.14 shows the accuracy of the top 10 features of digital camera 2 according to the P_{rank} indicating that five of the top ten features received more than 95% accuracy, namely, software, quality, picture, camera, and size.

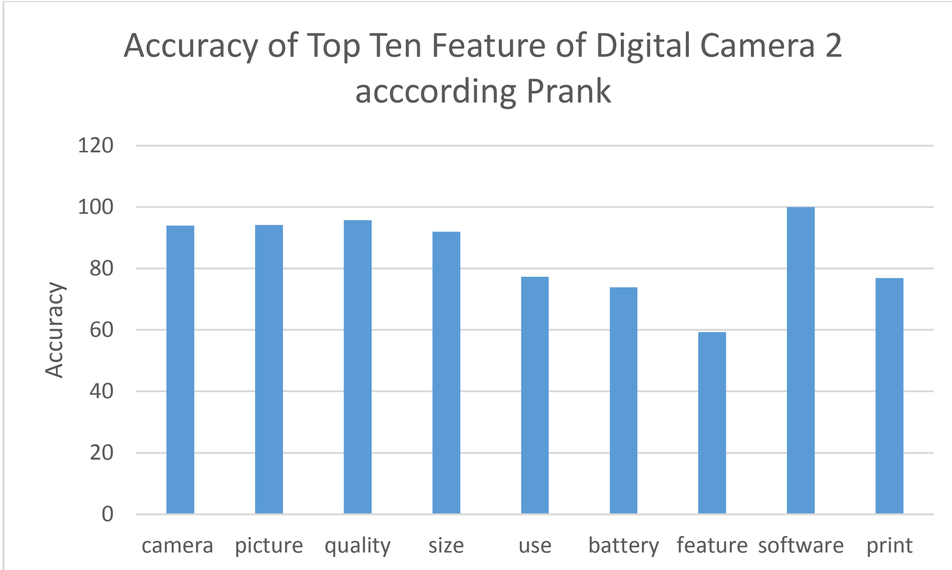


Figure 5.14: Accuracy of Top Ten Features of Digital Camera 2 according to P_{rank}

The features 'Use', 'Print' and 'Battery' having 78%, 77% and 74% accuracy, respectively. However, the 'Feature' has only 59% accuracy. The average accuracy of top 10 features of digital camera 2 according to P_{rank} was found to be 85%.

Figure 5.15 represents the top ten features of digital camera 2 according to the N_{rank} showing the shortcomings and limitations of the camera. The feature 'Picture' has the highest N_{rank} score indicating that many users are not satisfied by this feature. Few users showed dissatisfaction on the features, namely, 'Battery, 'Transfer', 'Autofocus', 'Movie', 'Camera', and 'Service'. However, the features 'Viewfinder', 'LCD' and 'Software' reported negatively by very few users.

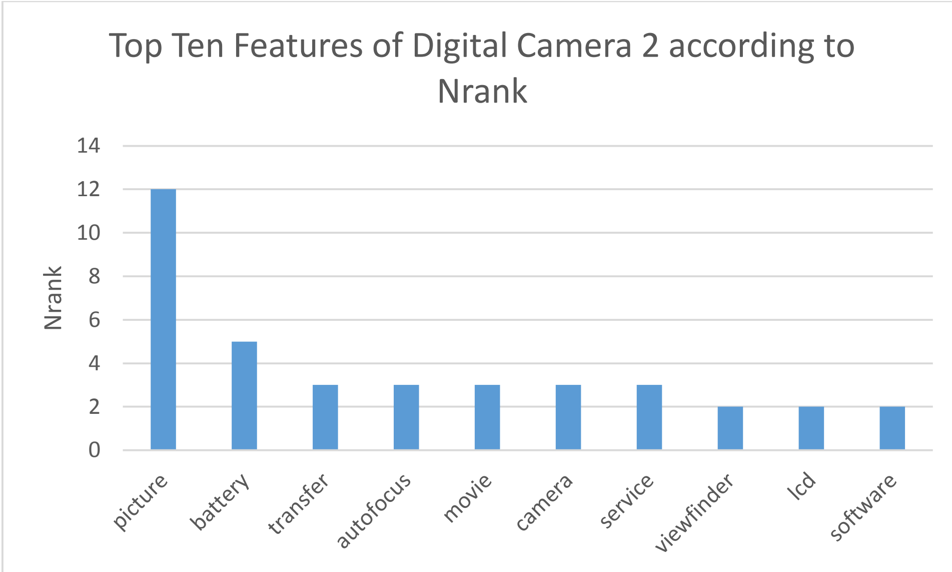


Figure 5.15: Top Ten Features of Digital Camera 2 according to N_{rank}

Figure 5.16 shows the accuracy of the top 10 features of digital camera 2 according to the N_{rank} indicating that nine of the top ten features received 100% accuracy, namely, battery, transfer, autofocus, movie, camera, service, viewfinder, LCD, and software. The last feature is ‘picture’ having 91% accuracy. The overall accuracy of the Opinion Analyzer according to N_{rank} was found to be 99%.

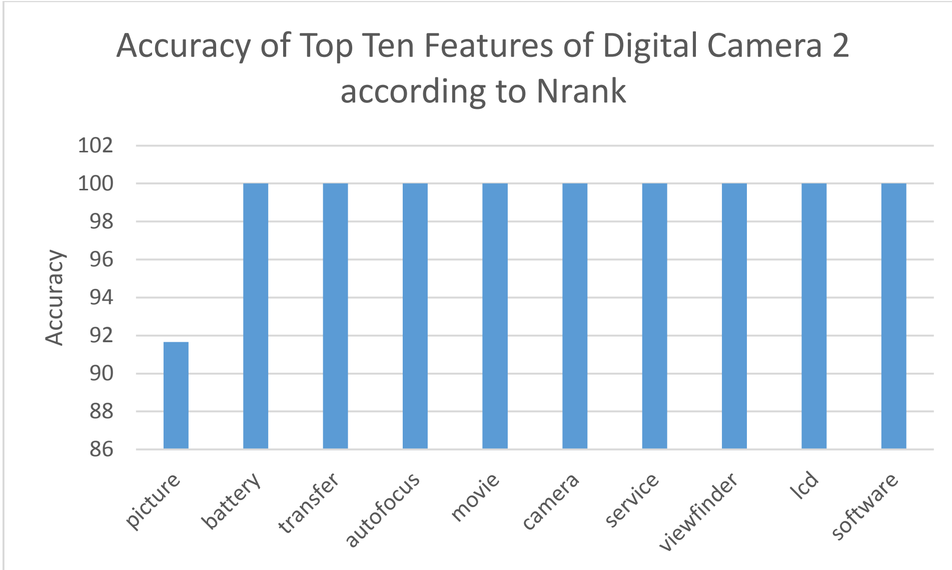


Figure 5.16: Accuracy of Top Ten Features of Digital Camera 2 according to N_{rank}

Figure 5.17 represents the top ten features of digital camera 2 according to the O_{rank}. The O_{rank} score of 145 received by the camera indicates that the camera is acknowledged by many users.

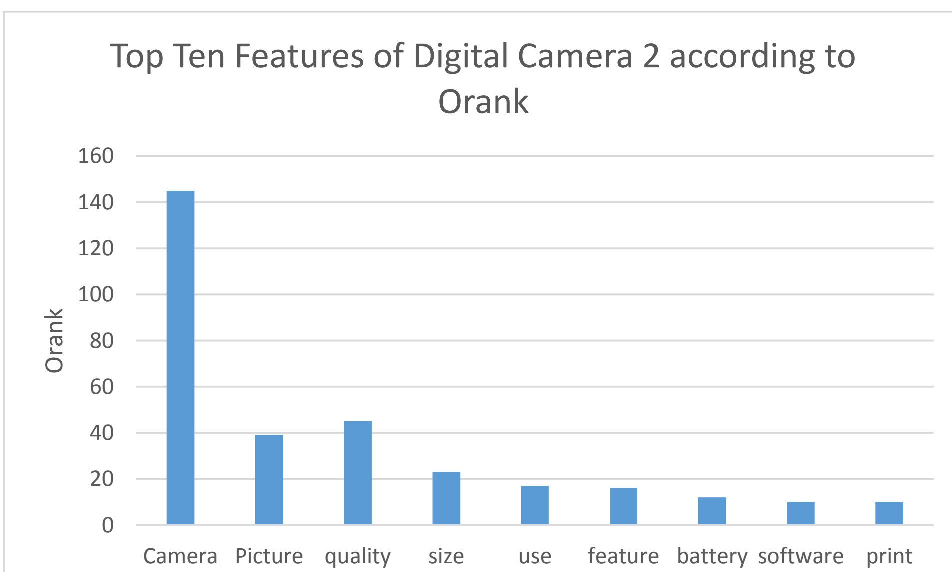


Figure 5.17: Top 10 Features of Digital Camera 2 according to O_{rank}

The O_{rank} score of the feature 'Picture' and 'Quality' also highlights customers' satisfaction about these features. The features 'Size', 'Use', 'Feature', 'Battery, 'Software' and 'Print' showing that users commented on these features overall positively.

Figure 5.18 shows the accuracy of the top 10 features of digital camera 2 according to the O_{rank} indicating that good accuracy is achieved by the Opinion Analyzer as four of the top ten features received around 100% accuracy, namely, software, picture, camera, and size.

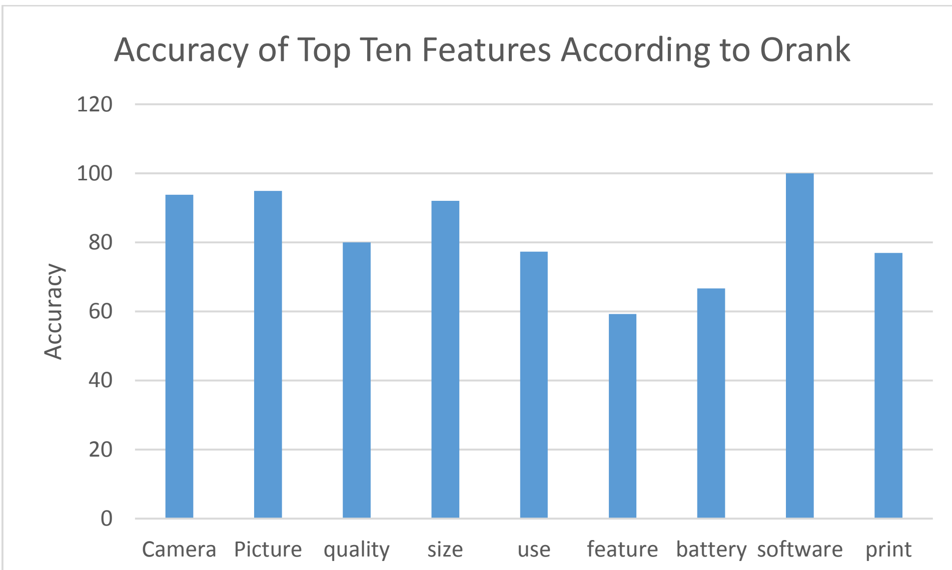


Figure 5.18: Accuracy of Top Ten Features of Digital Camera 2 according to O_{rank}.

The next three features are 'Quality', 'Use' and 'Print' having accuracy greater than 75%. However, the 'Feature' and 'Battery' features have accuracy less than 75%. The overall accuracy of the Opinion Analyzer according to O_{rank} is 74%.

It can be inferred from Figure 5.7 and 5.13 that the top three features according to the P_{rank} of these cameras are the same, namely, camera, picture and quality indicating that users discussed about these features in a highly positive manner. The results are in accordance with the results of Ahmad and Doja (2012) and Liu et al. (2005) who reported the feature 'Picture' as one of the top features according to positive comments. However, these cameras are different in terms of their limitations. For instance, the viewfinder, software and camera are the weak points of the digital camera 1 whereas the shortcomings of the digital camera 2 are picture, battery and transfer as indicated by Figure 5.9 and 5.15. Liu et al. (2005) also described the features 'Picture' and 'Battery' of the digital camera 2 as one of the top features according to negative opinions.

5.1.2.3 Feature Ranking of Cellular Phone

Figure 5.19 presents the top ten features of 'Cellular Phone' according to the P_{rank} that highlights the strengths of the phone.

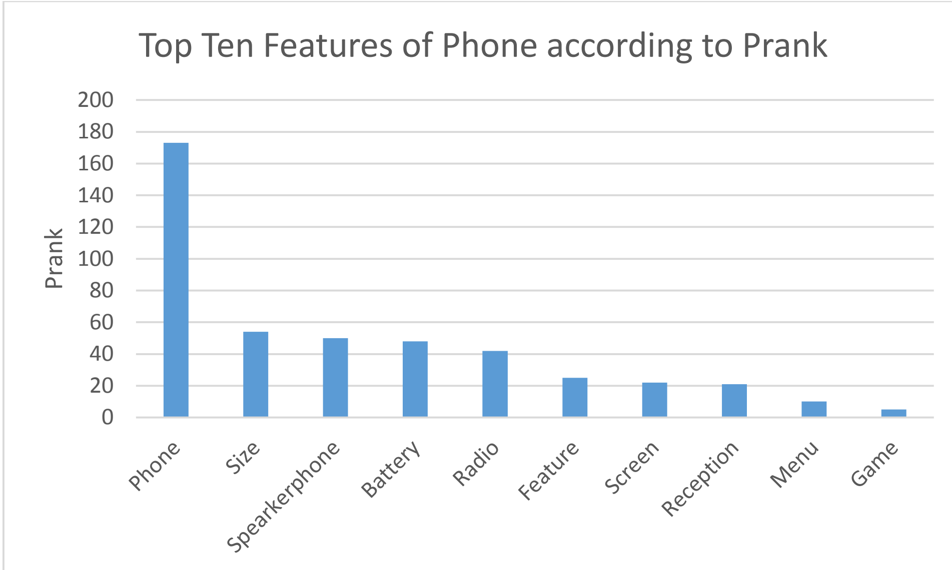


Figure 5.19: Top Ten Feature of Cellular Phone According to P_{rank}

The P_{rank} of the features 'Phone', 'Size' and 'Speakerphone' shows that a large number of customers appreciated these features. The other main features indicating the customers' satisfaction are 'Battery', 'Radio', 'Feature', 'Screen' and 'Reception' having P_{rank} in the range between 42 and 21. However, the 'Menu' and 'Game' features of the cellular phone are commented positively by very few users.

Figure 5.20 shows the accuracy of the top 10 features of the cellular phone according to the P_{rank} showing that notable accuracy is achieved by the Opinion Analyzer as four of the top ten features received 100% accuracy, namely, size, feature, menu, and game. The next four features are 'Speakerphone', 'Screen', 'Battery' and 'Radio' having accuracy greater than 85%. The features 'Reception' and 'Phone' have accuracy 75% and 72%, respectively, resulting in 90% of overall accuracy.

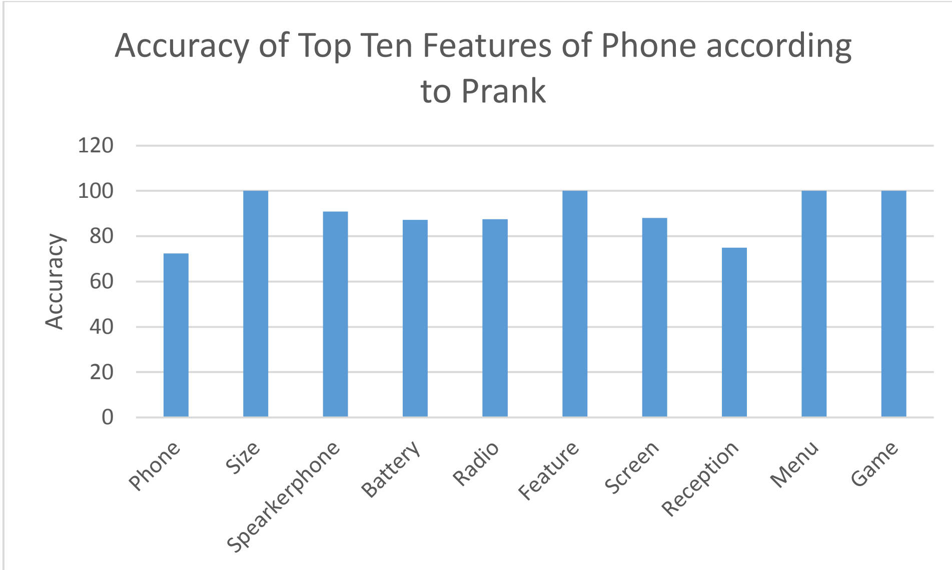


Figure 5.20: Accuracy of Top Ten Features of Cellular Phone according to P_{rank}

Figure 5.21 represents the top ten features of the cellular phone according to the N_{rank} indicating the limitations of the phone. The feature ‘Game’ has the highest N_{rank} score exhibiting that many users are not satisfied by this feature of the phone. The features ‘Screen’, ‘Internet’, ‘Radio’, and ‘Camera’ also received customers’ dissatisfaction. The features ‘Tone’, ‘Color’, ‘Menu’, ‘Reception’ and ‘Sound’ have N_{rank} equals to 3 representing that very few users reported these feature negatively.

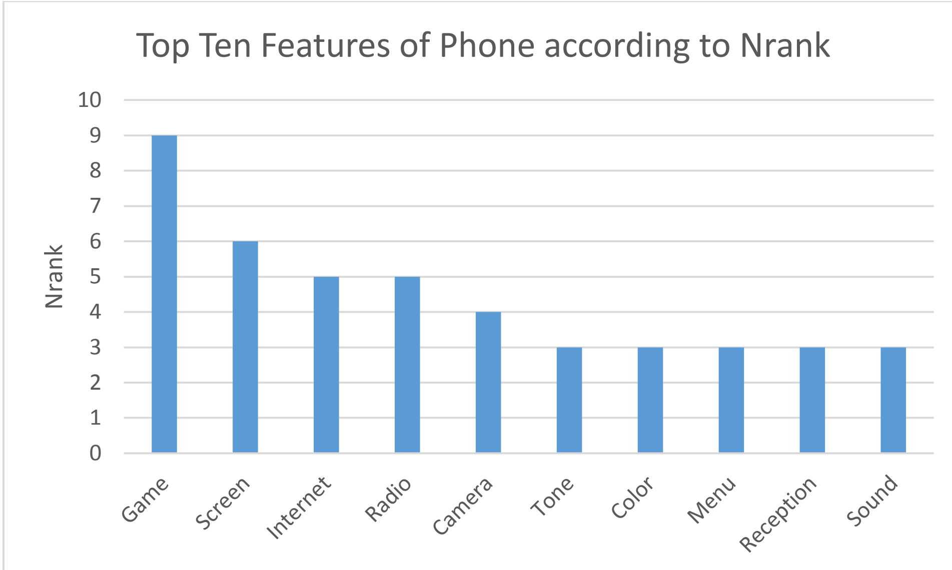


Figure 5.21: Top Ten Feature of Cellular Phone According to N_{rank}

Figure 5.22 shows the accuracy of the top 10 features of cellular phone according to the N_{rank}. Eight of the top ten features received 100% accuracy, namely, game, screen, internet, radio, camera, tone, color, menu, reception, and sound. The remaining two features are 'Internet' and 'Menu' having accuracy of 71% and 60%, respectively. The accuracy of the top ten features according to the N_{rank} was found to be 93%.

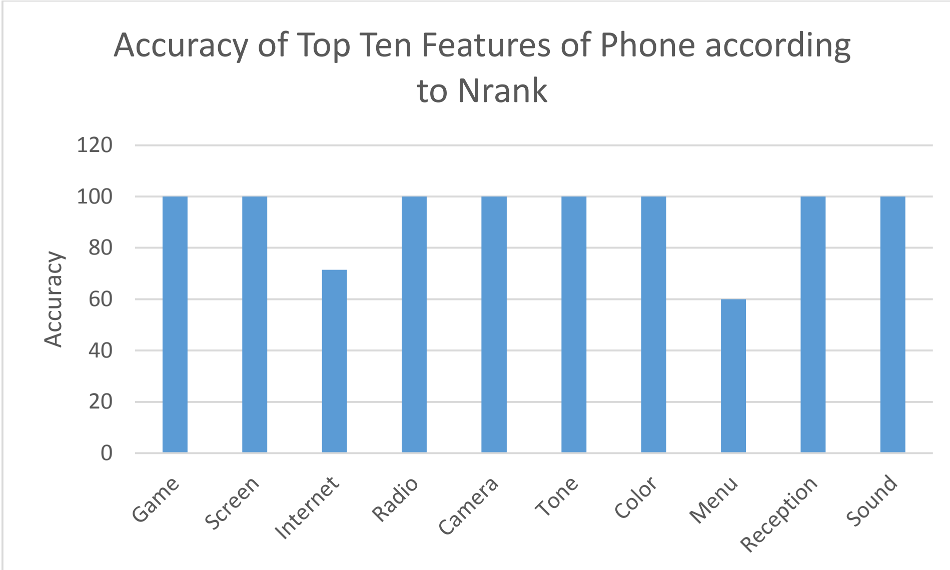


Figure 5.22: Accuracy of Top Ten Features of Cellular Phone according to N_{rank}

Figure 5.23 represents the top ten features of the cellular phone according to the O_{rank} highlighting the overall score of the phone.

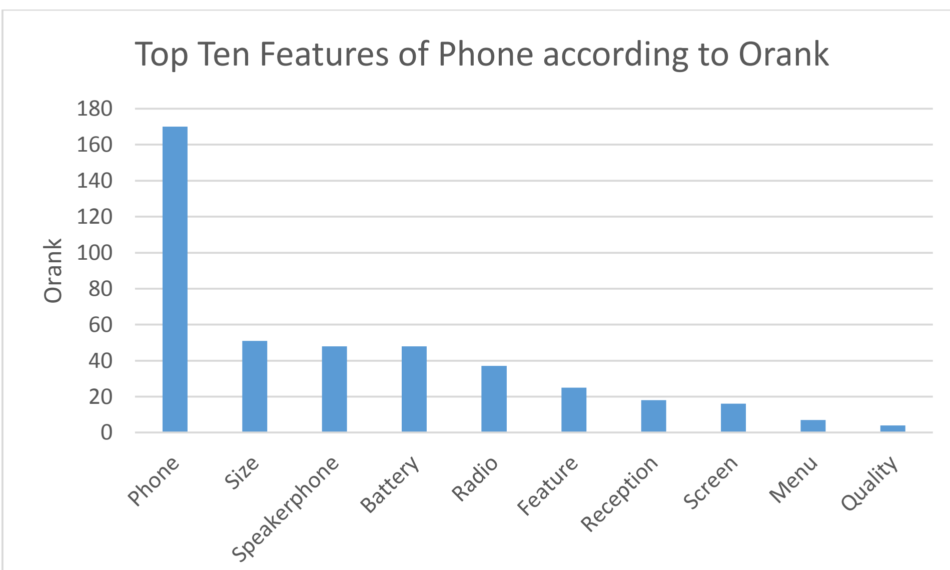


Figure 5.23: Top Ten Features of Cellular Phone according to O_{rank}

Overall, customers acknowledged the phone to a large extend as indicated by its O_{rank}. The feature ‘Size’, ‘Speakerphone’, ‘Battery’ and ‘Radio’ have O_{rank} score between 51 and 37, respectively, reflecting that customers’ are satisfied with these features of the phone to a large extend. Users also discussed the features ‘Feature’, ‘Reception’, ‘Screen’, ‘Menu’ and ‘Quality’ positively.

Figure 5.24 shows the accuracy of the top 10 features of cellular phone according to the O_{rank}. Three of the top ten features received 100% accuracy, namely, size, feature and quality. The next three features ‘Speakerphone’, ‘Battery’ and ‘Radio’ having accuracy greater than 85%. The features ‘Reception’ and ‘Menu’ have accuracy of 72% and 71%, respectively. However, the feature ‘Screen’ has accuracy of 12.5% only. Therefore, the overall accuracy of the Opinion Analyzer according to O_{rank} dropped to 79%.

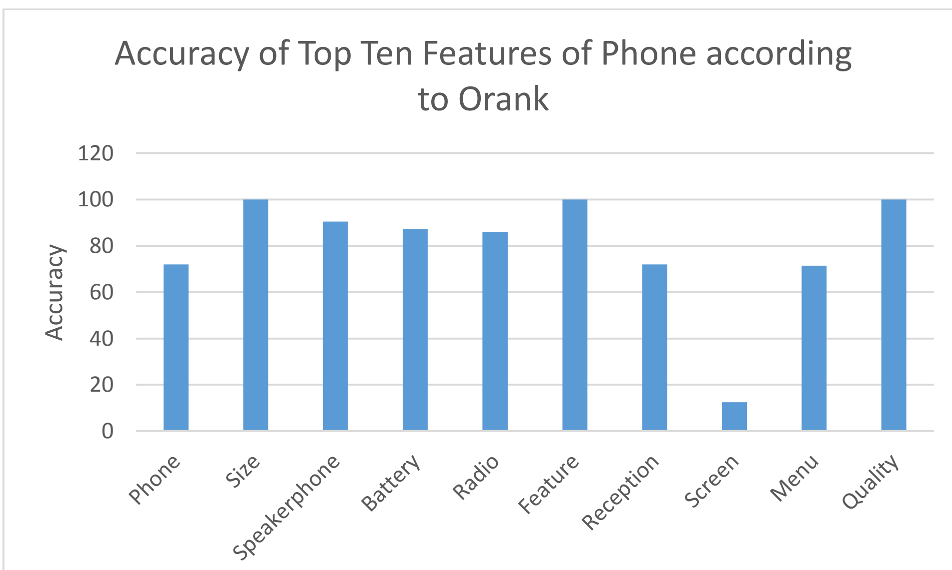


Figure 5.24: Accuracy of Top Ten Features of Phone according to O_{rank}

5.1.2.4 Feature Ranking of MP3 Player

Figure 5.25 represents the top ten features of 'MP3 Player' according to the P_{rank} showing the features of the 'MP2 Player' appreciated by users. The feature 'Player' overwhelmingly received positive feedback from a number of users as its P_{rank} is 226. The other top features include 'Software', 'Sound' 'Battery' and 'Price' also acknowledge by a large number of users having P_{rank} 114, 107, 103 and 92, respectively. The other features showing customers satisfaction are 'Product', 'Playlist', 'Case', 'Button' and 'Unit'.

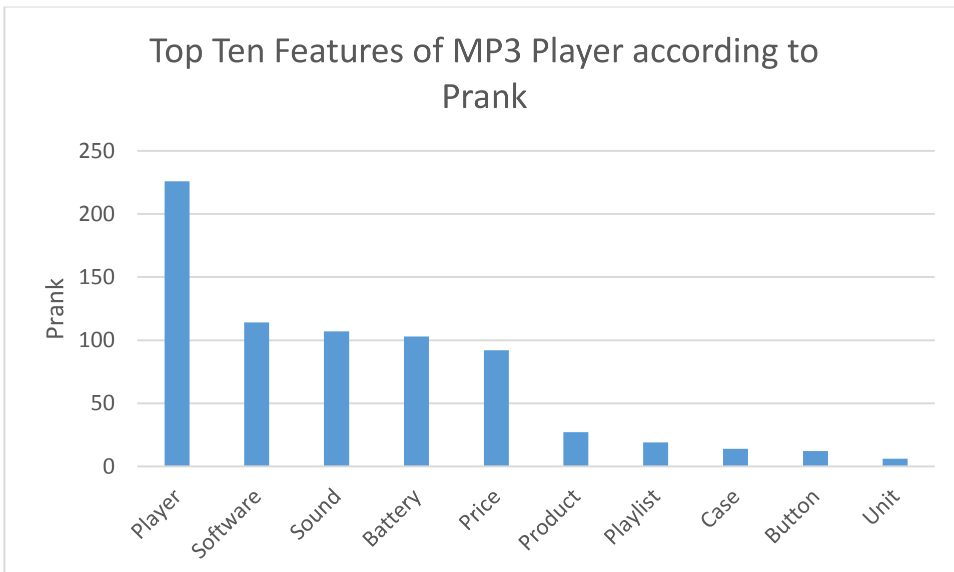


Figure 5.25: Top 10 Features of MP3 Player according to P_{rank}

Figure 5.26 shows the accuracy of the top 10 features of MP3 Player according to P_{rank}. As, the accuracy of three of the top ten features, namely, 'Product', 'Unit', and 'Service' is 100% indicating that noteworthy accuracy is achieved by the Opinion Analyzer. Further, there are four other features with accuracy greater that 85% that include 'Software' (93%), 'Player' (86%), 'Sound' (89%), and 'Battery' (91%). The features 'Price', 'Product', 'Button' have accuracy in the range between 80% and 77. The overall accuracy achieved by the Opinion Analyzer is 89%.

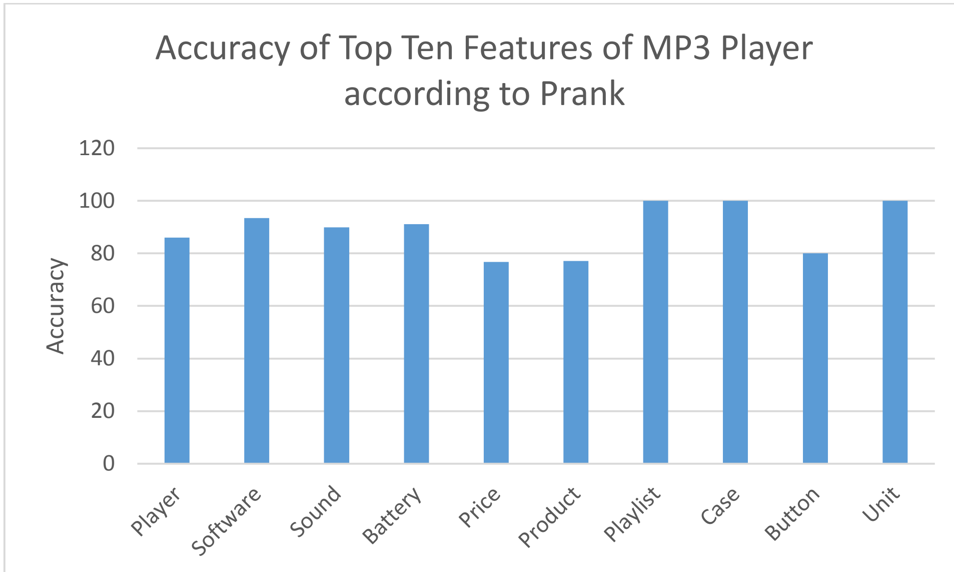


Figure 5.26: Accuracy of Top Ten Features of MP3 Player according to P_{rank}

Figure 5.27 displays the top ten features of MP3 Player according to the N_{rank} that refers to the shortfalls of the discussed product.

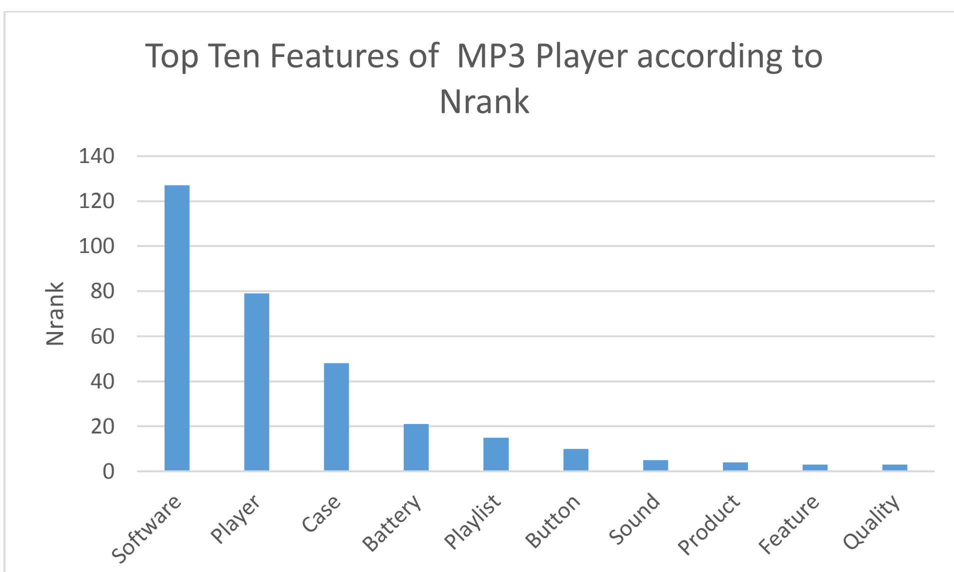


Figure 5.27: Top 10 Features of MP3 Player according to N_{rank}

A majority of users negatively discussed the feature 'Software' of the MP3 Player as the N_{rank} of the feature is 127. The features 'Player', and 'Case' are the other features having relatively larger N_{rank} score of 79 and 48, respectively, showing customers' disappointment. The features 'Battery', 'Playlist', and 'Button' having N_{rank} scores of 21, 15 and 10, respectively, signifying that some users are not satisfied by these features. The feature 'Sound', 'Product', 'Feature', and 'Quality' are also discussed negatively by very few users.

Figure 5.28 shows the accuracy of the top 10 features of MP3 Player according to the N_{rank}.

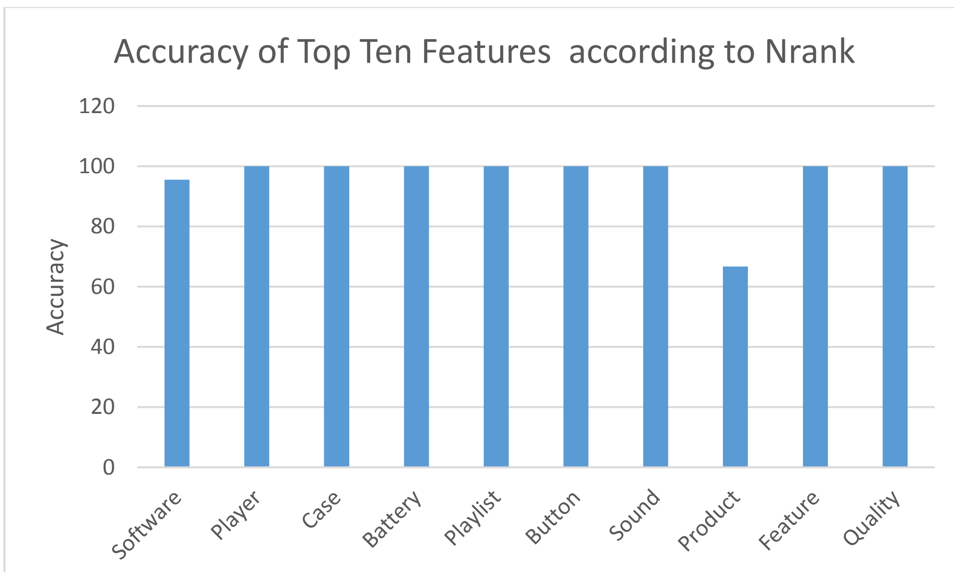


Figure 5.28: Accuracy of Top Ten Features according to N_{rank}

The Opinion Analyzer achieves very good accuracy as eight of the top ten features achieved 100% accuracy, namely, player, case, battery, playlist, button, sound, feature and quality. The feature 'Software' is also up to the mark as its accuracy is 95%. However, the feature

‘Product’ only achieved 67% accuracy. The Opinion Analyzer achieved an overall accuracy of 96%.

Figure 5.29 pinpoints the top ten features of MP3 player according to the O_{rank} by emphasizing the overall score of the product and features.

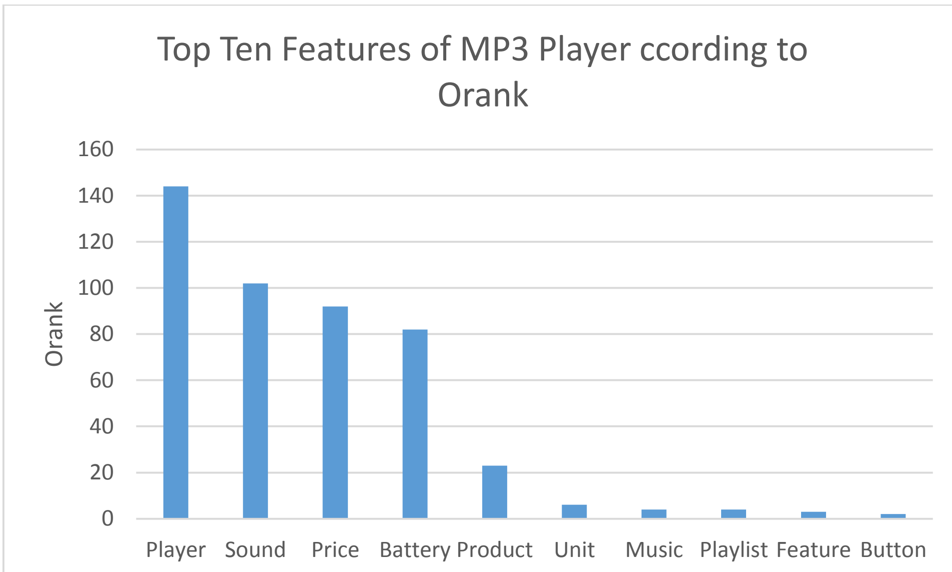


Figure 5.29: Top 10 Features of MP3 Player according to O_{rank}

The player received O_{rank} score of 144 indicating that a large number of customers showed satisfaction on the player. The feature ‘Sound’, ‘Price’, and ‘Battery’ also commented positively by many users. The O_{rank} of feature ‘Product’ shows that users positively concerned about this feature. The other important features are ‘Unit’, ‘Music’, ‘Playlist’, ‘Feature’, and ‘Button’ having O_{rank} score in a close range between 6 and 2.

Figure 5.30 shows the accuracy of the top 10 features of MP3 Player according to the O_{rank} indicating that satisfactory accuracy is achieved by the Opinion Analyzer as three of the top ten features achieved 100% accuracy, namely, unit, playlist and feature).

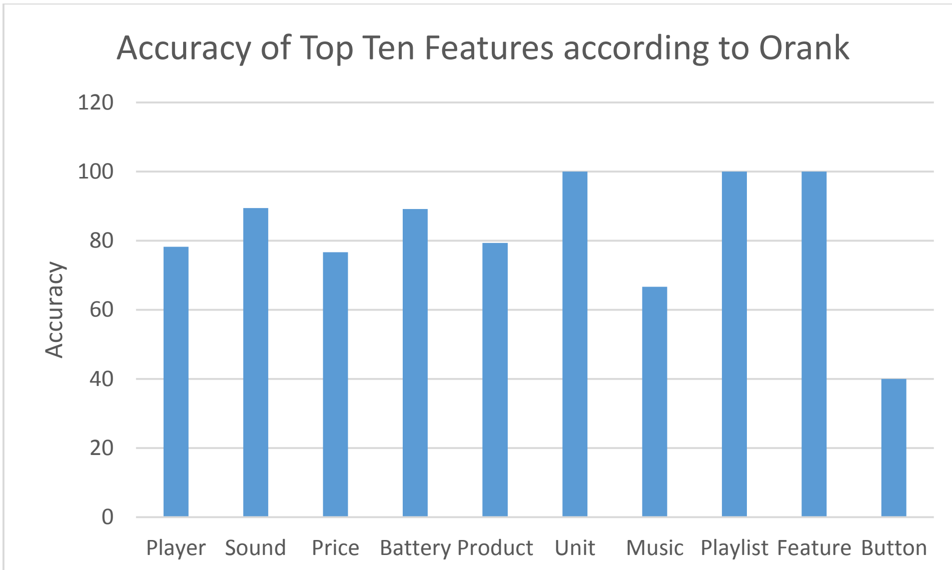


Figure 5.30: Accuracy of Top Ten Features according to O_{rank}

The next two features are 'Sound' and 'Battery' having approximately 90% accuracy, followed by the features 'player', 'price', 'product' and music have accuracy in the range between 80% and 65%. However, the accuracy of the feature 'button' is only 40%. Irrespective of the fact that one of the features has 40% accuracy, even then the overall accuracy achieved by the Opinion Analyzer is 82%.

5.1.2.5 Feature Ranking of DVD Player

Figure 5.31 highlights the top ten features of 'DVD Player' according to the P_{rank} describing the stong points of the product. The feature 'Player' is positively appreciated by a number of users because it has highest P_{rank}. The other top features include 'Play', 'Feature' and 'Price' having P_{rank} 21, 23 and 28, respectively, indicating positive customers'opinions on these features. The other features which are appreciated by a few number of users include 'Apex', 'Picture', 'Work', 'Product' and 'Unit'.

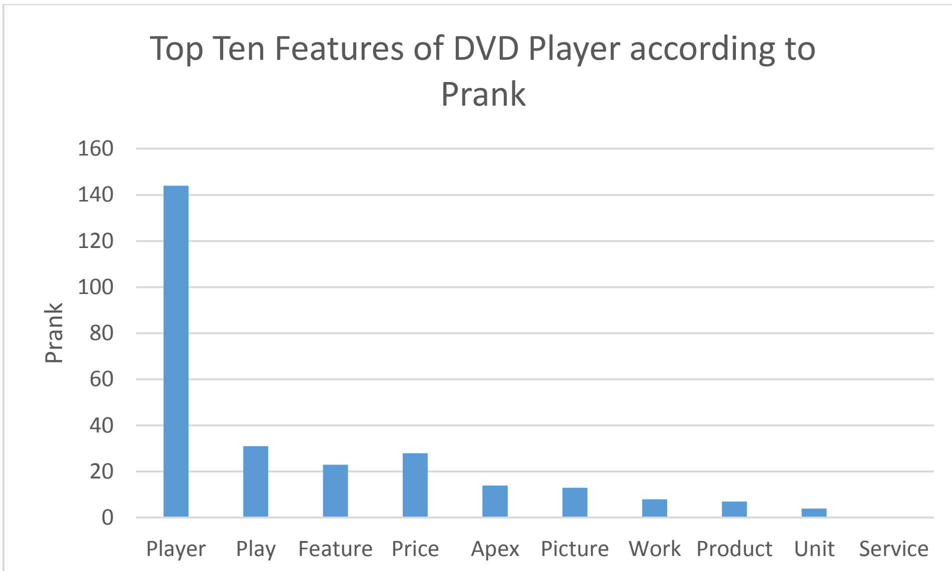


Figure 5.31: Top 10 Features of DVD Player according to P_{rank}

Figure 5.32 shows the accuracy of the top 10 features of DVD player according to the P_{rank}. The accuracy of four of the top ten features that include 'Product', 'Unit', 'Service', and 'Feature' is 100% indicating that notable accuracy is achieved by the Opinion Analyzer. Further, there are four other features having accuracy above 85% that includes 'Player' (86%), 'Play' (90%), 'Apex' (92%), and 'Work' (87.5%). The feature 'Picture' has an accuracy of 76%. However, the feature 'Feature' has accuracy of only 60%. In spite of it, the overall accuracy achieved by the Opinion Analyzer is 90%.

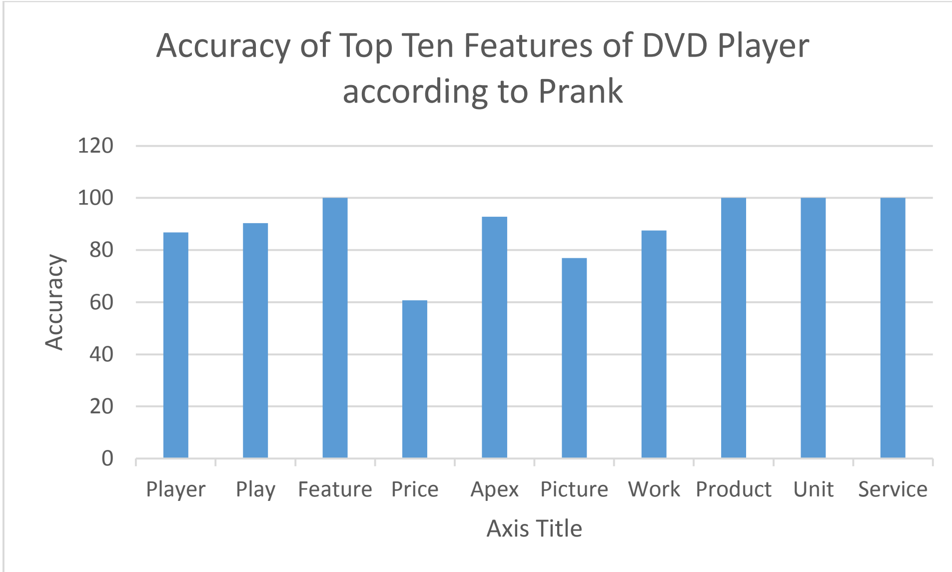


Figure 5.32: Accuracy of Top 10 Features of DVD Player according to P_{rank}

Figure 5.33 represents the top ten features of DVD player according to the N_{rank} that refers to the inadequacies of the DVD player.

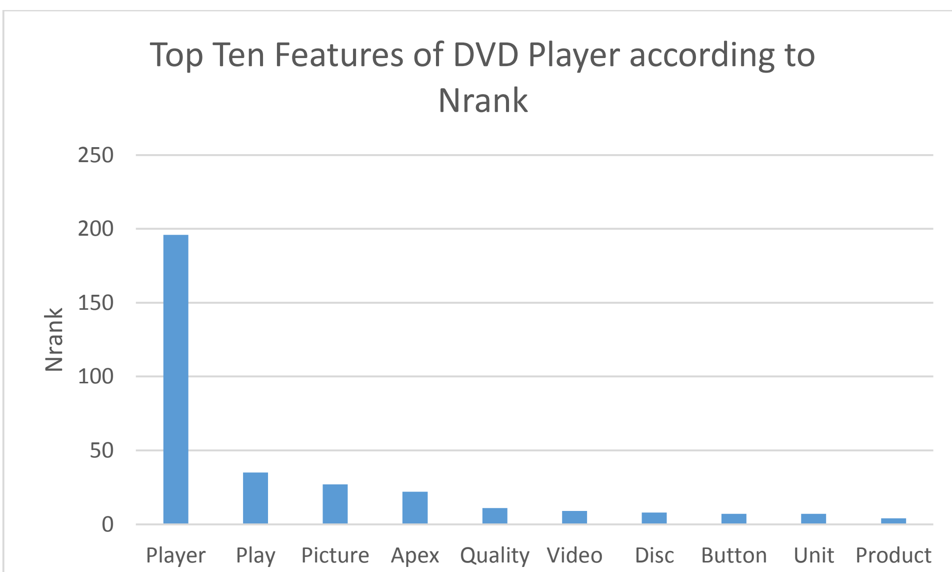


Figure 5.33: Top 10 Features of DVD Player according to N_{rank}

A majority of users negatively discussed the DVD player as indicated by its $N_{rank,}$ that is, 196. The feature ‘Play’, ‘Picture’, ‘Apex’ and ‘Quality’ are the other features having relatively larger N_{rank} in the range between 35 and 10 showing customers’ disapproval. The next five features are ‘Video’, ‘Disc’, ‘Button’, ‘Unit’ and ‘Product’ indicating that a few number of users are not satisfied by these features of DVD player.

Figure 5.34 shows the accuracy of the top 10 features of DVD Player according to the N_{rank}. The Opinion Analyzer achieved good accuracy as three of the top ten features achieved 100% accuracy, namely, apex, button and unit. There are three features that received more than 80% accuracy that include ‘Player’, ‘Play’ and ‘Product’. The other three features that achieved more than 60% accuracy are ‘Picture’, ‘Video’ and ‘Disk’. However, there is one feature named ‘Quality’ whose accuracy is only 58% resulting in overall accuracy of 81%.

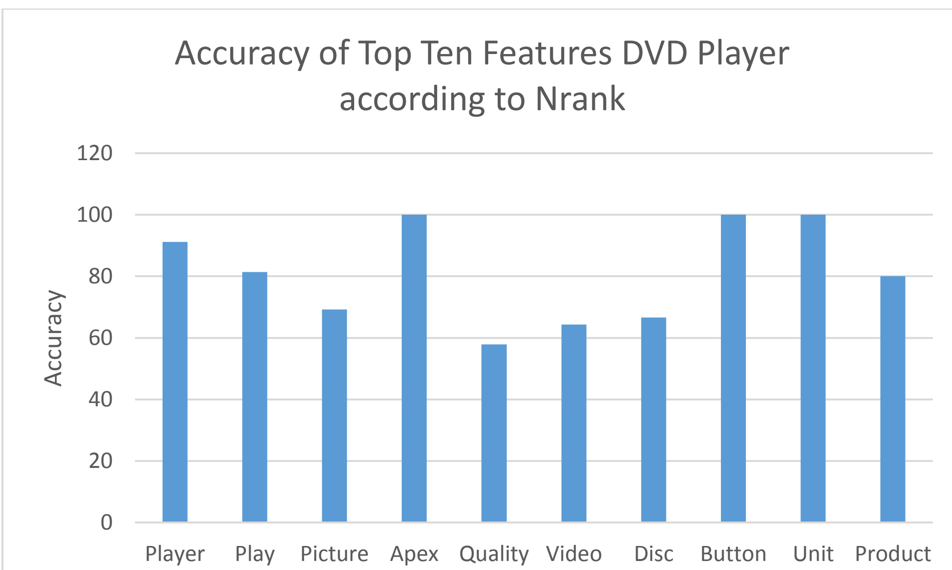


Figure 5.34: Accuracy of Top 10 Features of DVD Player according to N_{rank}

Figure 5.35 highlights the top ten features of DVD player according to the O_{rank} by emphasizing the overall score of the DVD player and its features. The top four features has positive O_{rank} depicting that users are relatively satisfied by these features that include 'Feature' (23), 'Price' (17), 'Work' (7), 'Product' (3). However, the O_{rank} score of six of the features is negative illustrating that these features of DVD player are unable to satisfy majority of users. The features having negative O_{rank} are 'Unit', 'Service', 'Play', 'Button', 'Disc' and 'Apex' having O_{rank} score of -3,-4, -7, -7, -8 and -9, respectively.

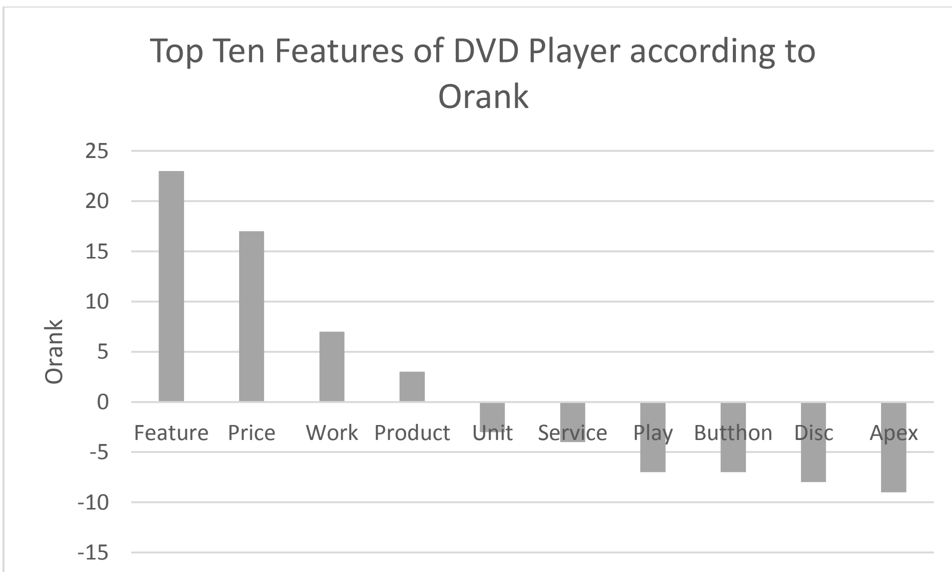


Figure 5.35: Top 10 Features of DVD Player according to O_{rank}

Figure 5.36 shows the accuracy of the top 10 features of DVD player according to the O_{rank} indicating that satisfactory accuracy is achieved by the Opinion Analyzer as four of the top ten features achieved 100% accuracy, namely, feature, unit, service and button.

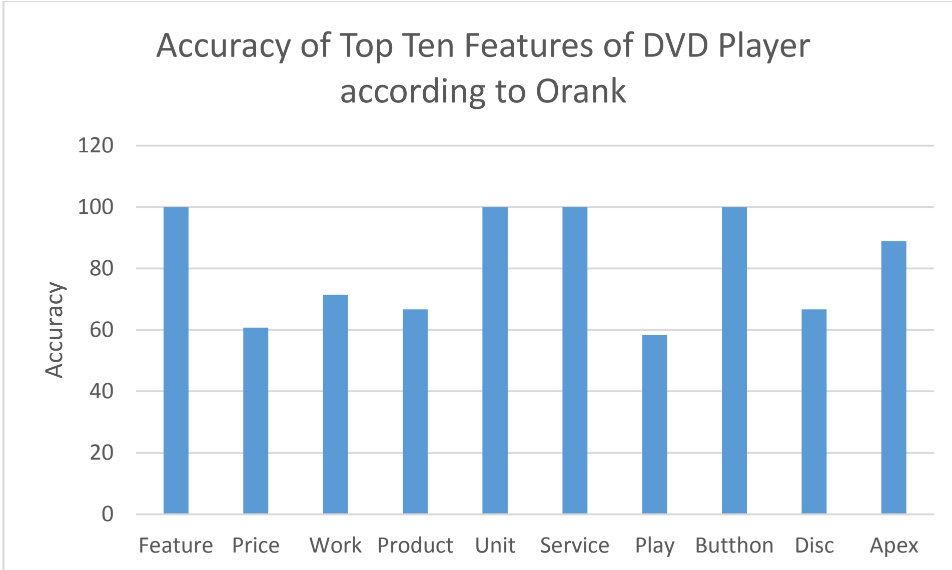


Figure 5.36: Accuracy of Top 10 Features of DVD Player according to O_{rank}

The next four features in line according to accuracy of the Opinion Analyzer are ‘Apex’, ‘Work’, ‘Product’ and ‘Disc’ having accuracy greater than 65%. However, the features ‘Price’ and ‘Play’ have accuracy less than or equal to 60%. Irrespective of the fact that one of the features have accuracy less than 60%, even then the overall accuracy achieved by the Opinion Analyzer is 80%.

The Opinion Analyzer investigates critical features of all the products, which have significant impact on future sales, reputation management, decision making, risk management, new products’ design, marketing strategies, and product adoption. The Opinion Analyzer can have several implications for entrepreneurs as well as for customers. The first implication of the proposed system from a business perspective is the identification of influential product features that have a significant impact on sales.

Online consumers’ reviews present tremendous challenges for marketers. One of the challenges is to develop interactive marketing practices for making connections with target consumers that capitalizes consumer-to-consumer communications to generate product adoption. Therefore, the second implication is for entrepreneurs who seek to venture consumers-to-consumers communications for advertisement and marketing campaigns. They can advertise the strenghts of their products (top features acording to P_{rank}) in their marketing campaigns to build product adoption and reputation, create awareness, trigger interest and generate sales. Viral marketing is the most intriguing strategy among consumer-leveraging possibilities by utilizing customer-to-costumer communications to disseminate information about a product or service using Internet. Futher, the Opinion Analyzer can be augmented in marketing campaigns by enterprises to promote and advertise their products that might result in incrased future sale.

Thirdly, the Opinion Analyzer empowers businesses to be more aware of the consensus surrounding their business and can formulate actions to resolve negative word-of-mouth and swing consumers’ opinions in their favor. For instance, the Opinion Analyzer identifies that the viewfinder of the digital camera 1 received all negative comments. Therefore, the future sale of the camera can be increased by improving its viewfinder. Similarly, the Opinion Analyzer suggests manufacturers to address the weaknesses of products (top features acording to N_{rank}) in upcoming products that reshapes future products’ design and plans. Consequently, the improved product designs and better future plans resulted in increased future sale. The opinion analyzer also facilitates in competative intelligence in which entrepreneurs can compare their products with competitive products that facilitates in decision making and pinpoints potential risks from competitors. For instance, the digital camera 1 can be compared with the digital camera 2 in order to get competitive intelligence. Additionally, from a consumer point of view, it highlights the strengths and weaknesses of a product for making purchase decisions.

5.1.2.6 Average Accuracy of the Opinion Analyzer

Figure 5.37 compares the average accuracy of the P_{rank} and N_{rank} of five products with the accuracy of the FBS system (Hu and Liu, 2004). The FBS system is selected for the comparison as the data set of the Opinion Analyzer and the FBS is same. Further, both systems utilized frequency-based approach. The Opinion Analyzer outperformed FBS system on the accuracy of digital camera 1, cellular phone, MP3 Player and DVD Player. For digital camera 2, the Opinion Analyzer showed a little performance degradation, however, the Opinion Analyzer outclassed FBS on the basis of average accuracy.

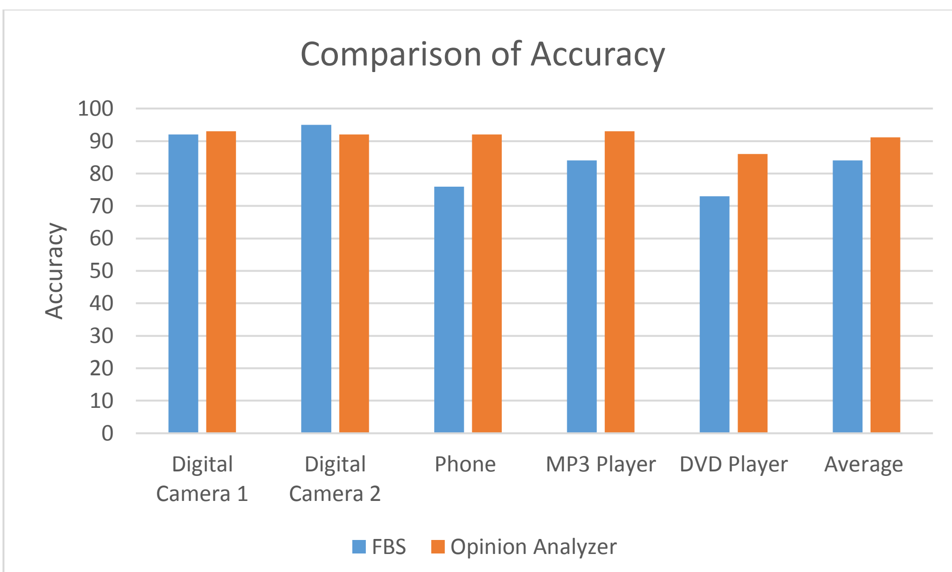


Figure 5.37: Accuracy of P_{rank}, N_{rank} and O_{rank}

5.2 Users' Preferences about Existing Opinion Visualizations

The visualizations were ranked in order to identify the users'preferences about existing opinion visualization. For the ranking of the opinion visualizations, descriptive statistics, such as mean and standard deviation of all metrics (discussed in Section 4.1.5.1) for each of the visualizations were analyzed. Mean + values were calculated for each metrics, that is, the average of the mean values of each metrics against all the visualizations. Table 5.5 shows the ranking of the visualizations. Contents in grey shade show the metrics which scored higher than the mean+ values. A visualization ranks higher than others based on the number of grey shaded metrics. The top three visualizations are bar chart, followed by glowing bar and tree map. Therefore, traditional bar chart metaphor was adopted to display feature ranking and tree map was modified to present opinion-strength-based summary as it is difficult to portray multi-dimension information using a bar chart.

Table 5.1: Ranking of the visualizations

Visualizations		Visual Appeal	Easy To Understand	User Friendly	Informativeness	Intuitiveness	Usefulness	Comprehensiveness	Comparison Ability	Representation Style	Pre-Knowledge Required
Mean+		3.08	3.08	3.03	3.19	2.77	3.09	2.99	2.95	3.01	2.94
BC	M	3.77	4.02	3.84	3.91	3.24	3.78	3.51	3.53	3.60	2.67
	SD	1.19	1.10	1.24	1.04	1.19	1.10	1.21	1.17	1.15	1.44
GB	M	3.88	3.73	3.65	3.69	3.10	3.63	3.42	3.31	3.43	2.79
	SD	1.09	1.26	1.18	1.14	1.10	1.08	1.15	1.17	1.24	1.36
TM	M	3.53	3.45	3.52	3.60	3.10	3.36	3.26	3.16	3.41	2.92
	SD	1.33	1.28	1.16	1.13	1.12	1.21	1.24	1.19	1.24	1.28

PC	M	3.25	3.60	3.51	3.39	2.93	3.36	3.18	3.03	3.08	2.44
	SD	1.12	1.26	1.18	1.13	1.34	1.15	1.21	1.15	1.06	1.33
VS	M	3.11	2.91	2.86	3.10	2.73	3.02	3.01	2.95	2.98	3.01
	SD	1.18	1.17	1.18	1.03	1.15	1.08	1.02	1.13	1.23	1.31
CRM	M	3.10	3.02	2.84	2.97	2.71	3.01	2.97	2.94	2.95	2.99
	SD	1.11	1.12	1.16	1.09	1.06	1.11	1.07	1.16	1.25	1.27
RP	M	2.77	2.83	2.71	2.96	2.66	2.93	2.82	2.80	3.03	3.07
	SD	1.10	1.23	1.25	1.09	1.02	1.09	1.13	1.21	1.21	1.40
BCS	M	3.01	3.09	3.01	2.99	2.68	2.97	2.83	2.82	2.64	2.86
	SD	1.24	1.23	1.23	1.12	1.17	1.13	1.11	1.13	1.21	1.26
OW	M	2.60	2.25	2.32	2.79	2.62	2.51	2.73	2.61	2.77	3.53
	SD	1.18	1.36	1.20	1.11	1.14	1.07	1.12	1.21	1.21	1.57
PM	M	2.58	2.81	2.85	3.01	2.40	2.84	2.67	2.66	2.79	2.80
	SD	1.19	1.30	1.15	1.17	1.12	1.25	1.10	1.18	1.18	1.40

BC: Bar Chart	GB: Glowing Bars	TM: Tree Map
PC: Pie Chart	VS: Visual Summary	CRP: Comparative Relation Map
RP: Rose Plot	BCS:Bar Chart With Symbols	OW: Opinion Wheel
PM: Positioning Map		

5.3 Opinion-Strength-based Visualization

To achieve the reseach objective 3, the tree map of Gamon et. al. (2005) (Figure 3.3 in Section 3.4.2) was selected and modified to present opinion-strength-based feature-level summary since it was in the top three opinion visualization techniques according to the users' preferences (Table 5.5 in Section 5.2). Basically, two modifications were made on the tree map of Gamon et. al. (2005) in this research work. The first modification is the use of color scale to present different level of opinion strength whereas the second modification is the division of rectangles according to opinion orientation and strength. The modified tree map displays multi-dimensional information, i.e. top features, semantic at various levels, feature weights, and comparison between positive and negative opinions about a particular feature, simultaneously. The modified opinion-strength-based feature-level summaries for five products are discussed in this section.

5.3.1 Opinion Summary of Digital Camera 1

Figure 5.38 presents the proposed tree map visualization for 'Digital Camera 1'. The proposed tree map is divided into ten rectangles. Each feature is rendered as one rectangle in the visualization. The size of the rectangle indicates feature weight. Each rectangle is sub-divided into different regions based on semantic orientation and strength to represent positive and negative semantic at three levels: weakly positive, mildly positive, strongly positive, weakly negative, mildly negative, and strongly negative. Different shades of red and green colors are used to encode levels of opinion strengths as shown in Figure 5.39. In contrast to the tree map (Gamon et al., 2005), our approach enables a much more detailed insight to show the comparison of positive and negative comments at various levels of opinion strength.

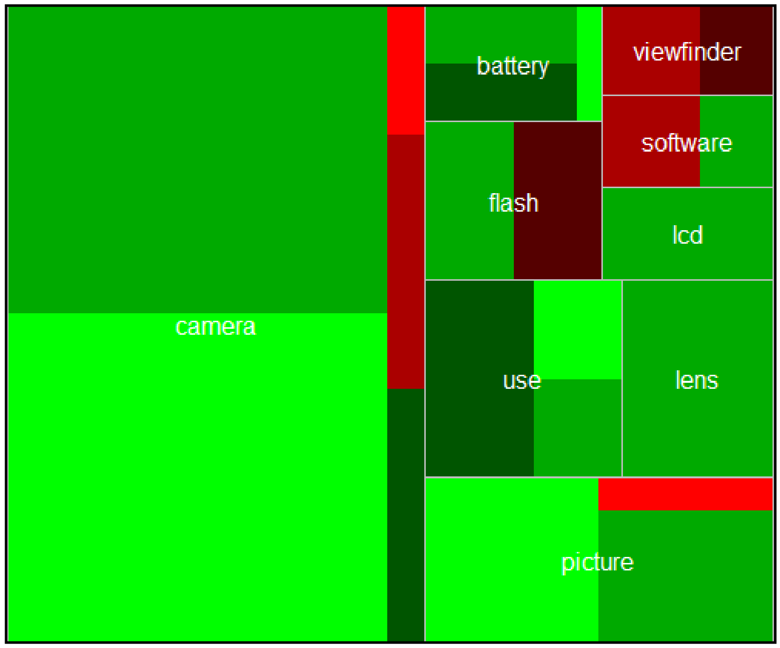


Figure 5.38: Proposed Tree Map Visualization showing Top Ten Features of Digital Camera 1

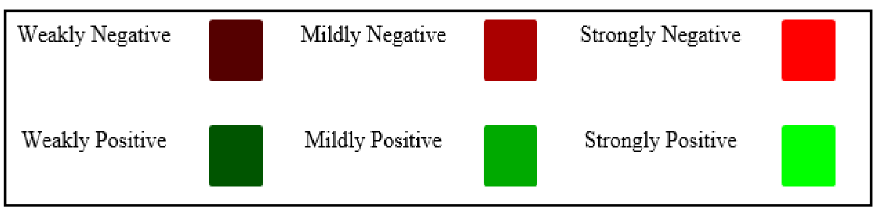


Figure 5.39: Color Scale

The size of the rectangle denotes the weight of the feature, therefore, it can be seen from Figure 5.38 that the feature 'Camera' has the highest weight illustrating that the highest number of users consider this feature. Most of the users discussed the feature 'camera' as strongly and mildly positive. Some of the users' comments about the 'camera' were weakly positive as well. However, the feature 'Camera' also received a few strongly negative and mildly negative comments from user. The other feature that is discussed by many users and is 'Picture' and received strongly positive and mildly positive comments. However, few users did not consider the 'Picture' feature in a positive way as their comments are strongly negative. The features 'Use' and 'Battery' received all positive comments including strongly, mildly and weakly positive opinions. It can be seen in Figure 5.38 that the feature 'Lens' and 'LCD' received all mildly positive comments. The feature 'Flash' of the digital camera 1 received both mildly positive and weakly negative comments and the feature 'Software' is addressed in mildly negative and mildly positive way by the users. However, the feature 'Viewfinder' received all negative comments as the comments it received are either weakly negative or mildly negative. The overall feedback of this product is visualized to be positive.

5.3.2 Opinion Summary of Digital Camera 2

Figure 5.40 displays the tree map visualization for 'Digital Camera 2' illustrating the visual summary along with the top features of the product according to their weights.

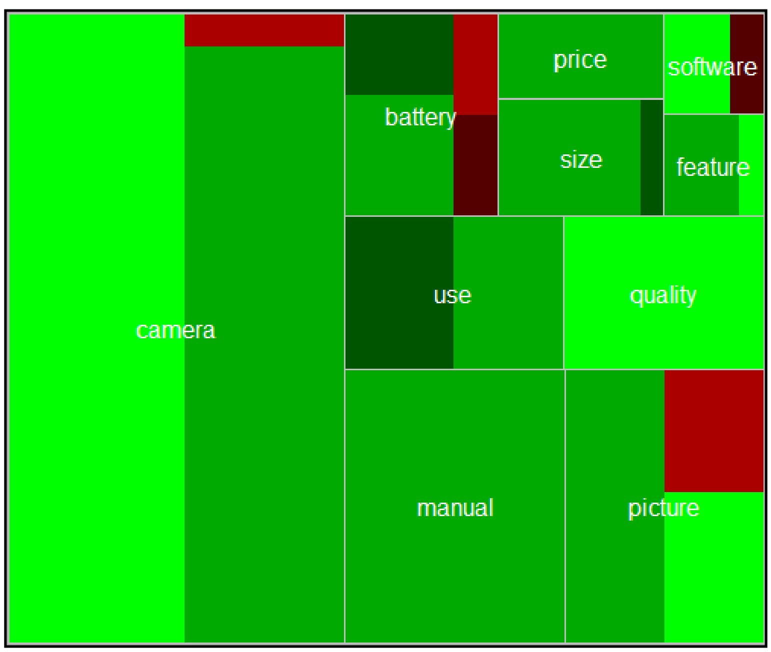


Figure 5.40: Proposed Tree Map Visualization showing Top Ten Features of Digital Camera 2

It can be seen that the feature ‘Camera’ has the highest weight similar to the product ‘Digital Camera 1’ representing most of users consider this feature in either strongly positive or mildly positive way. However, a few users also gave mildly negative opinions about the ‘Camera’. All of the users that discussed the features ‘manual’ and ‘price’ considered these features in mildly positive way. However, the feature ‘quality’ received all strongly positive comments describing that all of the users are satisfied by the quality of this camera to a larger extend. The other features that received all positive comments include ‘feature’ and ‘size’. Most of the users commented the feature ‘picture’ as positive except some of them discussed the ‘picture’ as a mildly negative. The features ‘battery’ and ‘software’ are considered positively by most of the users. However, according to some users the ‘battery’ and ‘software’ are negative features of this product. The overall feedback of this product seems to be positive.

From Figure 5.38 and 5.40, it can be deduced that the mostly discussed features in the digital camera domain are camera, picture, use, battery and software as these features are in the top ten features of both cameras. These features were considered by users most of the time while making a purchase decision. The results are in coordinace with Yang et al. (2010) who reported camera, picture, and battery as prominent features for the digital camera domain. Similarly, picutre and battery features were found to be significant features by Liu et al. (2005). According to ebay.com picture is a significant feature having an immense impact on future sale. Lee, Park and Ahn (2001) concluded a positive correlation between use of a product with its adoption and sale generation signalling that it is a vital feature. In accordance with the Lee et al. (2001), the feature ‘Use’ is also in mostly discussed feature.

5.3.3 Opinion Summary of Cellular Phone

Figure 5.41 shows the tree map visualization for ‘Cellular Phone’ depicting the visual summary that is mostly composed of green color indicating that most of the users addressed this product in a positive manner. The feature ‘Speakerphone’, ‘Size’, ‘Radio’ and ‘Screen’ are among the most widely discussed feature that received a majority of positive comments and a few number of negative comments. The features ‘Battery’, and ‘Reception’ received all positive comments that include strongly positive, weakly positive and mildly positive comments. The feature ‘Sound’ is the only feature in this product that is consider mildly positive by all of the users who addressed this feature. The overall feedback of this product seems to be positive.

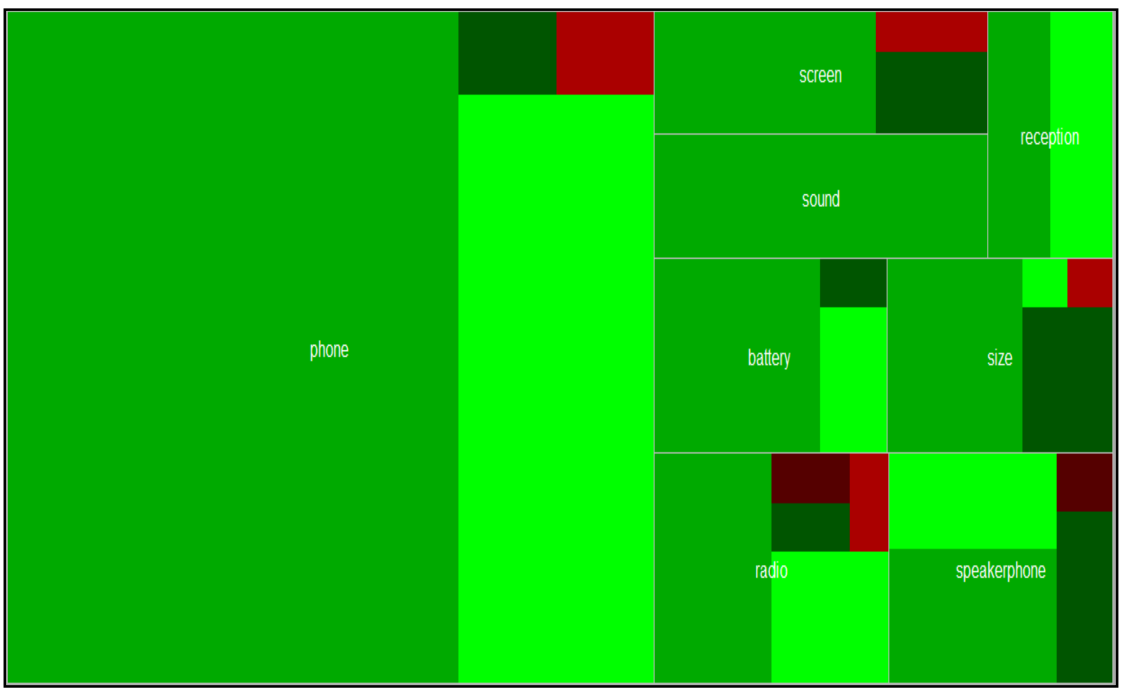


Figure 5.41: Proposed Tree Map Visualization showing Top Ten Features of Cellular Phone

5.3.4 Opinion Summary of MP3 Player

Figure 5.42 shows the tree map visualization for the product ‘MP3 Player’ illustrating the visual summary along with the weight of each of the top features of the product.

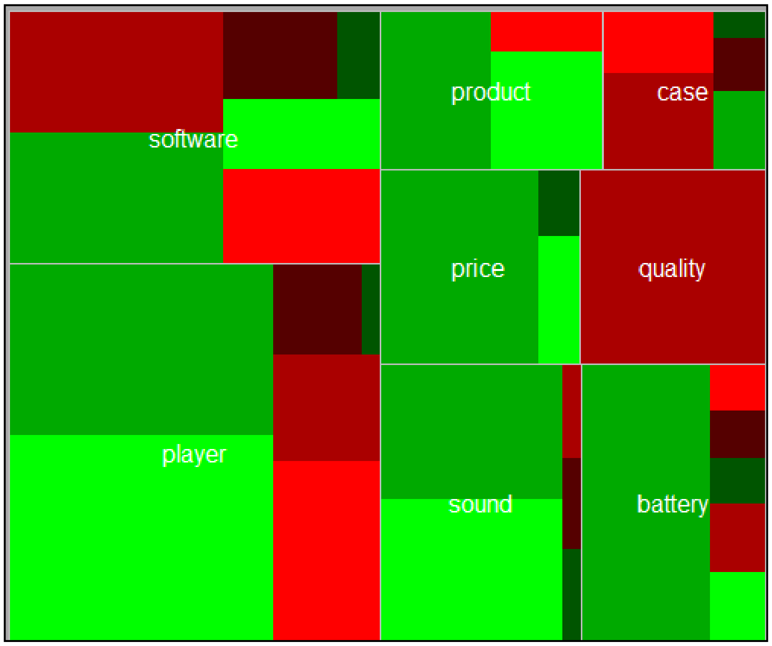


Figure 5.42: Proposed Tree Map Visualization showing Top Ten Features of MP3 Player

It can be seen from Figure 5.42 that the ‘Player’ has the highest weight representing that most of users consider this feature in positive way. However, some of the users also gave negative comments about the ‘Player’. The next most weighted feature is ‘Software’ that received a combination of both positive and negative comments. The features ‘Battery’, ‘Sound’, ‘Product’ receive more positive opinions as compared to negative ones. The feature ‘Quality is the only feature in the product that is considered mildly negative by all of the users. The users commented on the feature ‘case’ in a negative way, however very few user consider this feature as positive. The ‘Price’ is the only feature that received all positive comments. The overall feedback of this seems to be mixed (neither positive nor negative).

5.3.5 Opinion Summary of DVD Player

Figure 5.43 exhibits the tree map visualization for the product 'DVD Player' depicting the visual summary that is mostly composed of red color which indicates that most of the users addressed this product in a negative way.

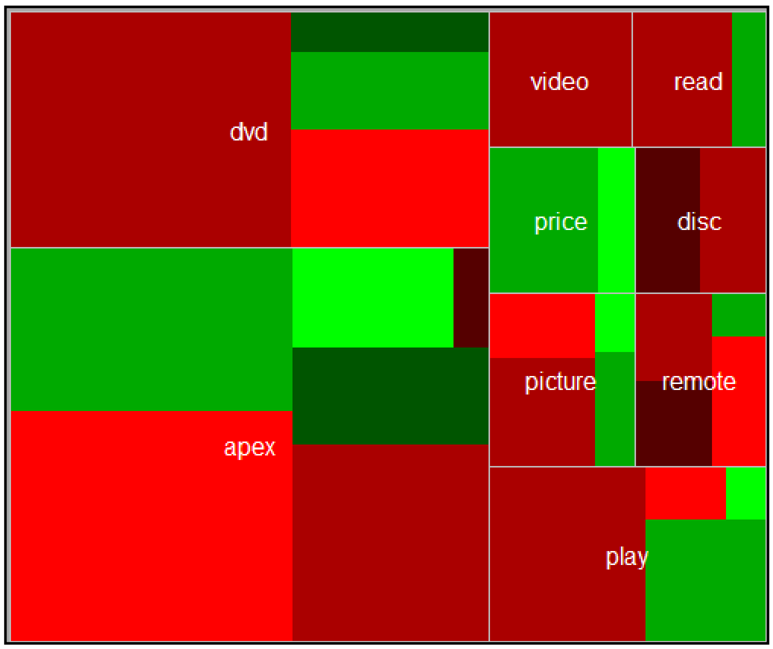


Figure 5.43: Proposed Tree Map Visualization showing Top Ten Features of DVD Player

The most discussed feature of the product is 'Apex' which received a majority of strongly and mildly negative comments. However, the 'Apex' also received some strong, mildly, and weak positive opinion. The second most discussed feature was 'dvd' that again received an overwhelming negative comments comprising of strong and mild negative comments. The third most discussed feature is 'Play' that also received more negative comments than the positive ones. The other features that received more negative comments as compared to positive comments are ‘Picture’, ‘Remote’, ‘Disc’ and ‘Read’. The feature ‘Video’ received all the mildly negative comments. However, the ‘Price’ feature is the only feature of the camera that received all positive comments depicting that besides the low quality of available the product, the price of the product is relatively less and is in reach of users. The overall feedback of this product is visualized to be negative.

5.4 Case Study

All participants were agreed and strongly agreed on the visual appeal of the visualization. Six participants reported agreement on understandability. Four participants reported agreement, whereas three participants showed strong agreement on user-friendliness. A majority of participants (6 participants) was agreed, while three participants described a strong agreement on the intuitiveness of the proposed visualization. Similar results were found for informativeness. Two participants suggested to define color scale with the visualization that will increase the understability of the visualization. We incorporated this suggestion (see Figure 5.44). A majority of participants highlighted that the width of the borders should be increased. We also amalgamated this suggestion. Another modification based on participants’ comments is the increase of font size.

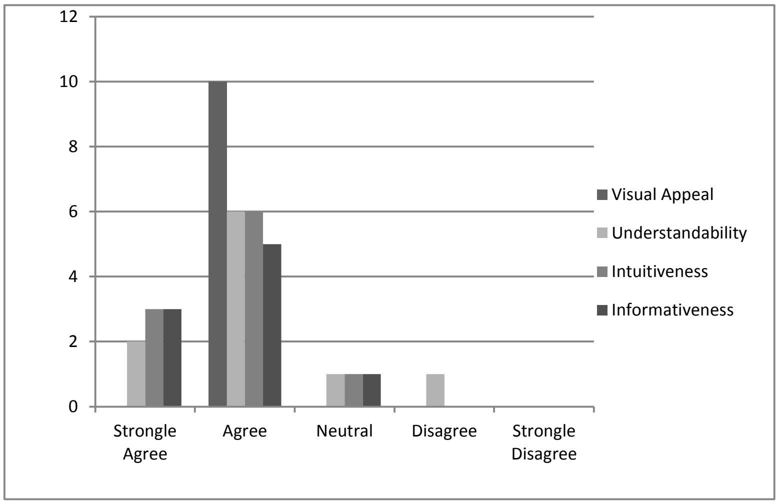


Figure 5.44: Result of Usability Study

Chapter 6 : Conclusion, Limitations and Future Work

This chapter concludes the research, and presents the limitations and possible future directions of this work.

6.1 Conclusion

The increase use and ubiquity of the Internet facilitates the dissemination of word-of-mouth through blogs, online forums, newsgroups, and consumers' reviews. E-WOM has become a powerful source of information for custumers and businesses that gauge customers' purchase intentions and enterprises' strategies. Besides, e-WOM influences customers' product adoption, satisfaction, trust and loyalty. However, it is difficult to analyze and summarize a large volume of e-WOM to obtain decision-oriented information that results in a research area called opinion mining. Feature-based opinion mining is a sub-area of opinion mining that attracted a great deal of researchers' attention recently aiming to extract features and corresponding semantics to ease the process of decision making for customers and businesses. The focus of this thesis was feature-based opinion mining due to its significant role in the opinion mining and wide range of applications such as business intelligence, recommender systems, news and citation analysis.

Online reviews vary greatly in quality, therefore it has become imperative to identify high quality reviews to enhance the decision making process. Although some review quality evaluation approaches are discussed in the literature to identify high quality review, however, their focus was not on the users' preferences. Moreover, most of the existing opinion mining systems ignore the quality of reviews. Additionally, current feature ranking approaches overlook opinion strength. Furthermore, existing opinion visualizations present overall positive and negative evaluation on each feature and are unable to exhibit opinion-strength-based summary.

This thesis is dedicated to integrate high quality review in feature ranking method with opinion-strength-based visualization to provide users with high quality decision-oriented information from enormous reviews. The objectives of the thesis are to (i) identify way(s) to incorporate users' preferences in ranking reviews, (ii) to enhance feature ranking using opinion strength, and (iii) to design an opinion-strength-based opinion visualization based on the users' preferences. In order to achieve these objectives, an opinion mining system called Opinion Analyzer was developed to incorporates high quality reviews for ranking critical products' features according to opinion orientation and strength. Furthermore, a questionnaire survey was conducted to obtain the users' preferences about existing opinion visualizations. The data were collected by conducting seminars and using a web-based online questionnaire (N=146). The collected data were analyzed using descriptive statistics. An opinion-strength-based visualization was designed based on the results of the survey and a usability study was conducted to access the usefulness of the proposed visualization.

First, in the Opinion Analyzer, the problem of selecting a set of high quality informative reviews according to the users' preferences was addressed. Second, a new feature ranking approach was proposed based on opinion orientation and strength that highlights the strengths and weaknesses of a target product. Third, an opinion-strength-based visualization was introduced that allows a detailed insight into products' features and corresponding sentiments at different levels of opinion strength. The first research objective was achieved by integrating users' preferences in the proposed review ranking method. The use of opinion strength in the proposed feature ranking method result in the acheivement of the second research objective. Third research objective was achieved by the development of the proposed opinion-strength-based visualization. Finally, the effectiveness of the proposed review and feature ranking methods were conducted in terms of accuracy and a user study was conducted to evaluate the efficacy of the proposed opinion-strength-based visualization, that is the last objective of the work.

The evaluation of the Opinion Analyzer was carried out on a real dataset of 332 reviews of five electronic products, namely, Canon PowerShot G3 camera, Nikon Coolpix 4300 Camera, Nokia 6610 Cellular Phone, Creative Labs Nomad Jukebox Zen Xtra 40GB MP3 Player and Apex AD2600 Progressive-scan DVD player. The Opinion Analyzer achieved promising accuracy for the proposed review quality evaluation and feature ranking for all the electronic products. Moreover, the Opinion Analyzer outperformed the FBS system in terms of accuracy. Furthermore, the result of the usability study conducted to access the usablilty of the proposed visualization suggested that the proposed opinion-strength-based visualization facilitates comparison between positive and negative opinions at different levels of opinion strengths of the top features with a high visual appeal, understandability, intuitiveness and informativeness.

Unlike the existing studies in this field, this work performs a ranking of reviews based upon the users' preferences. Also, the opinion strength in feature ranking is utilized to provide more accurate feature ranking. Further, the introduction of opinion-strength-based opinion visualization highlighted the critical product features and facilitated the comparison between the positive and negative opinions of a particular feature on different levels of opinion strength. The contributions of Opinion Analyzer are (i) the integration of users preferences in review ranking, (ii) the use of opinion strength in feature ranking that provide more accurate opinion summary, and (iii) the development of opinion-strength-based visualization.

6.2 Limitation and Future Work

Some final remarks are concerned with potential limitations of the study that suggests many promising directions for future research in the field of reviews quality classification and feature ranking. This section discusses limitations of the study along with future directions on which the current work can be improved further.

i. The proposed review ranking method is based on metadata and semantic features. Literature suggested that social features that are related to the reviewers characteristic such as reviewer rank, reviewer past reviews are helpful in predicting quality of reviews, therefore, future research should also incorporate social features in the proposed review quality method.

ii. The Opinion Analyzer utilizes frequency-based appraoch for the extraction of products' feature that tends to miss some low frequency features. For some users these features may have little importance, however, for other userrs the case is totally opposite. In future, a hybrid approach that assimilates both relation-based and frequency-based approach might be applied to extract low frequency features as well.

iii. Nouns are considered as products' features in this research work. However, noun phrases can be significant features of a target product. In future, the feature extraction task might be extended to extract both noun and noun phrases for the identification of vital products' features.

iv. Like, most of existing opinion mining systems, the Opinion Analyzer extracts only explicit features. However, a review may contains implicit features too. Future research should also try to extract and rank implicit features.

v. Future reseach work should exploit existing linguistic resources such as SentiWordNet to identify the semantic orientation of opinion words and estimate opinion strength to overcome the task of manual annotation in data set.

vi. This thesis reported the experimental results of electronic products, therefore, another future direction is to conduct experiments on reviews from different domains, such as book and cars reviews to evaluate the accuracy of the system on other domains.

vii. The Opinion Analyzer is tested with 332 reviews of five electronic products. In future a large data set comprising of numerous reviews of more products from different domains will be used to check the effectiveness of the Opinion Analyzer.

viii. The focus of the Opinion Analyzer is review format three (amazon.com). Future direction of this work is to consider the generalization of the Opinion Analyzer for other reviews formats such as review format one and two.

References

Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment Analysis in Multiple Languages : Feature Selection for Opinion Classification in Web Forums. *ACM Transactions on Information Systems (TOIS)*, *26*(3).

Ahmad, T., & Doja, M. N. (2012). Ranking System for Opinion Mining of Features from Review Documents. *IJCSI International Journal of Computer Science I*, *9*(4), pp. 440–447.

Andreevskaia, A., & Bergler, S. (2008). When Specialists and Generalists Work Together : Overcoming Domain Dependence in Sentiment Tagging. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL-2008)*,pp. 290–298.

Aue, A., & Gamon, M. (2005). Customizing Sentiment Classifiers to New Domains : a Case Study. In *Proceedings of Recent Advances in Natural Language Processing (RANLP-2005)*, pp. 1-5.

Baccianella, S., Esuli, A., & Sebastiani, F. (2009). Multi-facet Rating of Product Reviews. *LNCS 5478*, pp. 461–472.

Bai, X., White, D., & Sundaram, D. (2011). Purposeful Visualization. In *Proceedings of the 44th Hawaii International Conference on System Sciences - 2011*,pp. 1–10.

Bakhshi, S., Kanuparthy, P., & Shamma, D. A. (2015). Understanding Online Reviews : Funny , Cool or Useful ? In *CSCW*, pp. 1270–1276.

Bansal, M., Cardie, C., & Lee, L. (2008). The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. In *Proceedings of the International Conference on Computational Linguistics (COLING)*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.147.1132, pp. 15-18

Barbosa, L., & Feng, J. (2010). Robust Sentiment Detection on Twitter from Biased and Noisy Data. In *Proceedings of the International Conference on Computational Linguistics (COLING-2010)*,pp. 36–44.

Benamara, F., & Popescu, V. (2011). Towards Context-Based Subjectivity Analysis. In *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP-2011)*, pp. 1180–1188.

Binali, H., Potdar, V., & Wu, C. (2009). A State Of The Art Opinion Mining And Its Application Domains. In *Proceedings of the IEEE International Conference on Industrial Technology (ICIT '09)*. Monash University, Gippsland, pp 1-4.

Bjørkelund, E., Burnett, T. H., & Nørvåg, K. (2012). A study of opinion mining and visualization of hotel reviews. In *Proceedings of the 14th International Conference on*

Information Integration and Web-based Applications & Services - IIWAS '12 (p. 229). New York, New York, USA: ACM Press. doi:10.1145/2428736.2428773

Blair-goldensohn, Sasha, Hannan, K., Mcdonald, R., Neylon, T., Reis, G. A., & Reynar, J. (2008). Building a Sentiment Summarizer for Local Service Reviews. In *Proceedings of WWW-2008 workshop on NLP in the Information Explosion Era*.

Blitzer, J. (2006). Biographies , Bollywood , Boom-boxes and Blenders : Domain Adaptation for Sentiment Classification. In *In Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2007).*

Branavan, S. R. K., Chen, H., Eisenstein, J., & Barzilay, R. (2008). Learning Document-Level Semantic Properties from Free-text Annotations. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL-2008)*.

Brien, C. O. (2011). The Emergence of the Social Media Empowered Consumer. *Irish MarketingReview*.

Buche, A., Chandak, M. B., & Zadgaonkar, A. (2013). Opinion Mining and Analysis: A Survey. *International Journal on Natural Language Computing (IJNLC)*, *2*(3), 39–48.

Carenini, G., Ng, R., & Pauls, A. (2006). Multi-Document Summarization of Evaluative Text. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)* (pp. 305–312).

Chauchat, A. S. J., Eric, L., Lumière, U., & Mendès-france, P. (2008). Opinion Mining Issues and Agreement Identification in Forum Texts. In *Workshop data mining opinions (FODOP'08) Conjunction with the conference INFORSID* (pp. 51–58).

Chen, C. C., & Tseng, Y.-D. (2010). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, *50*(4), 755–768. doi:10.1016/j.dss.2010.08.023

Chen, C., Ibekwe-sanjuan, F., Sanjuan, E., & Weaver, C. (2006). Visual Analysis of Conflicting Opinions. In *IEEE Symposium On Visual Analytics And Technology (2006)* (pp. 59–66).

Cheng, X., & Xu, F. (2008). Fine-grained Opinion Topic and Polarity Identification. In B. M. J. M. J. O. S. P. D. T. Nicoletta Calzolari (Conference Chair) Khalid Choukri (Ed.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)* (pp. 2710–2714). Marrakech, Morocco: European Language Resources Association (ELRA).

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales : Online Book Reviews. *Journal of Marketing Research*, *43*(3), 345–354.

Choi, Y., & Cardie, C. (2010). Hierarchical Sequential Learning for Extracting Opinions and their Attributes. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL- 2010)* (pp. 269–274).

Conrad, J. G., & Leidner, J. L. (2008). Professional Credibility : Authority on the Web. In *WICOW08, 2008*.

Cui, H. (2006). Comparative Experiments on Sentiment Classification for Online Product Reviews. In *Proceedings of AAAI-2006.*

Dalal, M. K., & Zaveri, M. A. (2014). OpinionMining from Online User Reviews Using Fuzzy Linguistic Hedges. *Applied Computational Intelligence and Soft Computing*, *2014*(1).

Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., & Lee, L. (2009). How opinions are received by online communities: a case study on amazon.com helpfulness votes. In *Proceedings of the 18th international conference on World wide web* (pp. 141–150). New York, NY, USA: ACM. doi:10.1145/1526709.1526729

Dasgupta, S., & Ng, V. (2009). Mine the Easy , Classify the Hard : A Semi-Supervised Approach to Automatic Sentiment Classification. In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP (ACL-2009).*

Dave, K., Way, I., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519–528). New York, NY, USA: ACM. doi:http://doi.acm.org/10.1145/775152.775226

Dey, L., & Haque, S. K. M. (2008). Opinion Mining From Noisy Text Data. In *AND '08* (Vol. 00, pp. 83–90). Singapore.

Ding, X., Liu, B., Yu, P. S., & Street, S. M. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the international conference on Web search and web data mining* (pp. 231–240). New York, NY, USA: ACM. doi:http://doi.acm.org/10.1145/1341531.1341561

Draper, G. M., Livnat, Y., & Riesenfeld, R. F. (2009). A survey of radial methods for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, *15*(5), 759–76. doi:10.1109/TVCG.2009.23

Eirinaki, M., Pisal, S., & Singh, J. (2012). Feature-based opinion mining and ranking. *Journal of Computer and System Science*, *78*(4), 1175–1184. doi:10.1016/j.jcss.2011.10.007

El-Halees, A. (2013, September). OpinionMining. *Computer Standards & Interfaces*. doi:10.1016/j.csi.2012.10.004

Esuli, A., Sebastiani, F., Consiglio, I., Moruzzi, V. G., Pura, M., & Belzoni, V. G. B. (2005). Determining Term Subjectivity and Term Orientation for Opinion Mining. In *Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics* (Vol. 2, pp. 193–200). Trento, IT.

Etzioni, O., Cafarella, M., Downey, D., Popescu, A., Shaked, T., Soderland, S., … Yates, A. (2005). Unsupervised Named-Entity Extraction from the Web : An Experimental Study. *Artificial Intelligence*, *165*(1), 91–134.

Filho, P. B., Brun, C., & Rondeau, G. (2012). A Graphical User Interface for Feature-Based Opinion Mining. In *Proceedings of the NAACL-HLT 2012: Demonstration Session* (pp. 5–8). Montre´al, Canada.

Fiszman, M., Demner-fushman, D., Lang, F. M., Goetz, P., & Rindflesch, T. C. (2007). Interpreting Comparative Constructions in Biomedical Text Lister Hill National Center for Biomedical Communications. In *Proceedings of BioNLP* (pp. 137–144).

Funk, A., Li, Y., Saggion, H., Bontcheva, K., & Leibold, C. (2008). Opinion analysis for business intelligence applications. In *Proceedings of the first international workshop on Ontology-supported business intelligence*. New York, NY, USA: ACM. doi:http://doi.acm.org/10.1145/1452567.1452570

Gamon, M. (2004). Sentiment classification on customer feedback data : noisy data , large feature vectors , and the role of linguistic analysis. In *Proceedings of International Conference on Computational Linguistics (COLING-2004).*

Gamon, M., Aue, A., Corston-Oliver, S., & Ringger, E. (2005). Pulse: Mining customer opinions from free text. In *International Symposium on Intelligent Data Analysis* (pp. 121–132). Retrieved from http://www.springerlink.com/index/94q1nrhfc8a4e8tn.pdf

Gamon, M., Basu, S., Belenko, D., Fisher, D., Hurst, M., & K, A. C. (2008). BLEWS : Using Blogs to Provide Context for News Articles. In *Proceedings of the International Conference on Weblogs and Social Media*.

Ganesan, K., & Kim, H. D. (2008). Opinion Mining: A Short Tutorial. Retrieved from http://kavita-ganesan.com/opinion mining tutorial

Ganun, G., Kakodkar, Y., & Marian, A. (2013). Improving the quality of predictions using textual information in online user reviews. *Information Systems*, *38*, 1–15. doi:10.1016/j.is.2012.03.001

Gao, S., & Li, H. (2011). A cross-domain adaptation method for sentiment classification using probabilistic latent analysis. In *Proceedings of the 20th ACM international conference on Information and knowledge management - CIKM '11*. New York, New York, USA: ACM Press. doi:10.1145/2063576.2063728

Ghose, A., & Ipeirotis, P. G. (2007). Designing Novel Review Ranking Systems : Predicting the Usefulness and Impact of Reviews. In *ICEC '07 Proceedings of the ninth international conference on Electronic commerce* (pp. 303–309).

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the Helpfulness and Economic Impact of Product Reviews : Mining Text and Reviewer Characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498–1512.

Gregory, M. L., Chinchor, N., Whitney, P., Carter, R., Hetzler, E., & Turner, A. (2006). User-directed Sentiment Analysis : Visualizing the Affective Content of Documents. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text* (pp. 23–30). Stroudsburg, PA, USA.

Guo, H., Zhu, H., Guo, Z., Zhang, X., & Su, Z. (2009). Product feature categorization with multilevel latent semantic association. In *Proceeding of the 18th ACM conference on Information and knowledge management - CIKM '09* (p. 1087). New York, New York, USA: ACM Press. doi:10.1145/1645953.1646091

Guo, H., Zhu, H., Guo, Z., Zhang, X., & Su, Z. (2010). OpinionIt : A Text Mining System for Cross-lingual Opinion Analysis. In *CIKM'10* (pp. 1199–1208). oronto, Ontario, Canada.

Hai, Z., Chang, K., & Kim, J. (2011). Implicit Feature Identificatiovia Co-occurrence Association Rule Mining. *LNCS 6608*, 393–404.

Hao, M. C., Rohrdantz, C., Janetzko, H., Keim, D. a., Dayal, U., Haug, L. E., … Stoffel, F. (2013). Visual sentiment analysis of customer feedback streams using geo-temporal term associations. *Information Visualization*, *12*(3-4), 273–290. doi:10.1177/1473871613481691

Hao, M., Rohrdantz, C., Janetzko, H., Dayal, U., Keim, D. A., Haug, L., & Hsu, M. (2011). Visual Sentiment Analysis on Twitter Data Streams. In *IEEE VAST* (pp. 275–276).

Hatzivassiloglou, V., & Wiebe, J. M. (2000). Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the 18th conference on Computational linguistics - (COLING-2000)* (Vol. 1, pp. 299–305). Morristown, NJ, USA: Association for Computational Linguistics. doi:10.3115/990820.990864

Herold, A., Lemnitzer, L., & Berlin, B. (2012). *CLARIN-D User Guide*. Retrieved from http://media.dwds.de/clarin/userguide/userguide-1.0.1.pdf

Horrigan, J. B. (2008). *Online Shopping Internet users like the convenience but worry Findings*. Retrieved from http://www.pewinternet.org/~/media/Files/Reports/2008/PIP_Online Shopping.pdf.pdf

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the 19th national conference on Artifical intelligence* (pp. 168–177). New York, NY, USA: ACM. doi:http://doi.acm.org/10.1145/1014052.1014073

Jakob, N., & Gurevych, I. (2010a). Extracting Opinion Targets in a Single- and Cross-Domain Setting with Conditional Random Fields. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010)* (pp. 1035–1045).

Jakob, N., & Gurevych, I. (2010b). Using Anaphora Resolution to Improve Opinion Target Identification in Movie Reviews. In *Proceedings of the ACL 2010 Conference Short Papers, ACLShort '10* (pp. 263–268). Stroudsburg, PA, USA: Association for Computational Linguistics.

Jalilvand, M. R., & Samiei, N. (2012). The effect of electronic word of mouth on brand image and purchase intention: An empirical study in the automobile industry in Iran. *Marketing Intelligence & Planning*, *30*(4), 460–476. doi:10.1108/02634501211231946

Jiang, P., Zhang, C., Fu, H., Niu, Z., & Yang, Q. (2010). An Approach Based on Tree Kernels for Opinion Mining of Online Product Reviews. *IEEE International Conference on Data Mining*, 256–265. doi:10.1109/ICDM.2010.104

Jijkoun, V., Rijke, M. de, & Wouter Weerkamp. (2010). Generating Focused Topic-specific Sentiment Lexicons. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010)* (pp. 585–594).

Jin, W., & Ho, H. H. (2009). A Novel Lexicalized HMM-based Learning Framework for Web Opinion Mining. In *Proceedings of International Conference on Machine Learning (ICML-2009)*.

Jindal, N., & Liu, B. (2006a). Identifying Comparative Sentences in Text Documents. In *Proceedings of ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR-2006)*.

Jindal, N., & Liu, B. (2006b). Mining Comparative Sentences and Relations. In *Proceedings of National Conf. on Artificial Intelligence (AAAI-2006)*.

Jindal, N., & Liu, B. (2008). Opinion Spam and Analysis. In *Proceedings of the 2008 International Conference on Web Search and Data Mining* (pp. 219–230).

Jindal, N., Morgan, S., & Liu, B. (2010). Finding Unusual Review Patterns Using Unexpected Rules. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2010)*.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. doi:10.1016/j.bushor.2009.09.003

Keikha, M., & Crestani, F. (2010). Linguistic Aggregation Methods in Blog Retrieval. *Information Processing and Management*, *48*, 467–475.

Khan, K., Baharudin, B. B., Khan, A., & Malik, F. (2009). Mining Opinion from Text Documents : A Survey. In *3rd IEEE International Conference on Digital Ecosystems and Technologies* (pp. 217–222).

Kim, S., Pantel, P., Chklovski, T., & Pennacchiotti, M. (2006). Automatically Assessing Review Helpfulness. In *EMNLP '06 Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (pp. 423–430).

Kim, S., Rey, M., & Hovy, E. (2004). Determining the Sentiment of Opinions. In *Proceedings of Interntional Conference on Computational Linguistics (COLING-2004)*.

Kobayashi, N., Iida, R., Inui, K., & Matsumoto, Y. (2004). Opinion Mining on the Web by Extracting Subject-Aspect-Evaluation Relations. In *Proceedings of AAAI-CAAW'06*.

Kobayashi, N., Inui, K., & Matsumoto, Y. (2007). Opinion Mining from Web Documents: Extraction and Structurization. *Transactions of the Japanese Society for Artificial Intelligence*, *22*(3), 227–238. doi:10.1527/tjsai.22.227

Kongthon, A., Angkawattanawit, N., Sangkeettrakarn, C., Palingoon, P., & Haruechaiyasak, C. (2010). Using an Opinion Mining Approach to Exploit Web Content in Order to Improve Customer Relationship Management. In *Proceedings of PICMET '10 Technology Management for Global Economic Growth*. Portland, Oregon, USA.

Kongthon, A., Haruechaiyasak, C., Sangkeettrakarn, C., Palingoon, P., & Wunnasri, W. (2011). HotelOpinion : An Opinion Mining System on Hotel Reviews in Thailand. In *Proceedings of PICMET '11:* (pp. 1–6).

Koppel, M., & Shtrimberg, I. (2004). Good News or Bad News ? Let the Market Decide. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications* (pp. 86–88).

Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, *11*(3), 205–217. doi:10.1016/j.elerap.2011.10.003

Ku, L., Liang, Y. T., & Chen, H. (2006). Opinion Extraction , Summarization and Tracking in News and Blog Corpora. In *Proceedings of AAAI-CAAW'06* (pp. 100–107).

Kunpeng, Z., Narayanan, R., & Choudhary, A. (2010). Voice of the Customers : Mining Online Customer Reviews for Product Feature-based Ranking. In *Proceedings of the 3rd conference on Online social networks*. Berkeley, CA, USA.

Lee, D., Park, J., & Ahn, J. (2001). On the Explanation of Factors Affecting E-Commerce Adoption. In *Proceedings of the International Conference of Information Systems 2001* (pp. 109–120).

Lee, J., Park, D.-H., & Han, I. (2011). The different effects of online consumer reviews on consumers' purchase intentions depending on trust in online shopping malls: An advertising perspective. *Internet Research*, *21*(2), 187–206. doi:10.1108/10662241111123766

Lei, Z., Liu, B., Lim, S. H., & Eamonn, O.-S. (2010). Extracting and Ranking Product Features in Opinion Documents. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (pp. 1462–1470).

Li, F., Han, C., Huang, M., & Zhu, X. (2010). Structure-Aware Review Mining and Summarization. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING-2010)* (pp. 653–661).

Li, S., Chen, Z., & Tang, L. (2011). Exploiting Consumer Reviews for Product Feature Ranking Categories and Subject Descriptors. In *SWSM*.

Li, S., Wang, Z., Zhou, G., Yat, S., & Lee, M. (2011). Semi-Supervised Learning for Imbalanced Sentiment Classification. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-2011)* (pp. 1826–1831).

Li, X., Connexis, F. W., Liu, B., & Street, S. M. (2010). Distributional Similarity vs . PU Learning for Entity Set Expansion. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010).* (pp. 359–364).

Li, Y.-M., Liao, T.-F., & Lai, C.-Y. (2012). A social recommender mechanism for improving knowledge sharing in online forums. *Information Processing & Management*, *48*(5), 978–994. doi:10.1016/j.ipm.2011.10.004

Lim, E.-P., Nguyen, V.-A., Jindal, N., Liu, B., & Lauw, H. W. (2010). Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on Information and knowledge management - CIKM '10*. New York, New York, USA: ACM Press. doi:10.1145/1871437.1871557

Lin, C., Road, N. P., & Ex, E. (2009). Joint Sentiment / Topic Model for Sentiment Analysis. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2009)* (pp. 375–384).

Lipsman, A. (2007). *Online consumer-generated reviews have significant impact on offline purchase behavior*.

Liu, B. (2006). *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Springer.

Liu, B. (2010). *Sentiment Analysis and Subjectivity in Handbook of Natural Language Processing* (2nd Editio, pp. 1–38).

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan and Claypool publishers.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the Web. In *Proceedings of the 14th international conference on World Wide Web* (pp. 342–351). New York, USA: ACM. doi:http://doi.acm.org/10.1145/1060745.1060797

Liu, B., & Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. In *Mining Text Data* (pp. 1–49). Chapter 1: Kluwer Academic Publishers.

Liu, J., Cao, Y., Lin, C., Huang, Y., & Zhou, M. (2007). Low-Quality Product Review Detection in Opinion Summarization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* (pp. 334–342).

Lo, Y. W., & Potdar, V. (2009). A review of opinion mining and sentiment classification framework in social networks. In *3rd IEEE International Conference on Digital Ecosystems and Technologies* (pp. 396–401). Istanbul, TURKEY: Ieee. doi:10.1109/DEST.2009.5276705

Long, C., Zhang, J., & Zhut, X. (2010). A review selection approach for accurate feature rating estimation. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (pp. 766–774). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1944566.1944654

Lu, Y., Tsaparas, P., Ntoulas, A., & Polanyi, L. (2010). Exploiting social context for review quality prediction. In *Proceedings of the 19th international conference on World wide web - WWW '10*. New York, New York, USA: ACM Press. doi:10.1145/1772690.1772761

Lu, Y., Zhai, C., & Sundaresan, N. (2009). Rated aspect summarization of short comments. In *Proceedings of the 18th international conference on World wide web - WWW '09*. New York, New York, USA: ACM Press. doi:10.1145/1526709.1526728

Mackiewics, J., & Yeats, D. (2014). Product Review Users ' Perceptions of Review Quality : The Role of Credibility , Informativeness , and Readability. *IEEE Tansactions on Professional Communication*, *57*(4), 309–324.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. *2008*. Cambridge University Press. doi:10.4016/28481.01

Maynard, D. (2013). *Practical Opinion Mining for Social Media*.

Mei, Q., Ling, X., Wondra, M., Su, H., & Zhai, C. (2007). Topic Sentiment Mixture : Modeling Facets and Opinions in Weblogs. In *Proceedings of International Conference on World Wide Web (WWW-2007).*

Miao, Q., Li, Q., & Dai, R. (2009). AMAZING: A sentiment mining and retrieval system. *Expert Systems with Applications*, *36*(3), 7192–7198. doi:10.1016/j.eswa.2008.09.035

Moghaddam, S. A. (2013). *Aspect-based Opinion Mining in Online Reviews*. SIMON FRASER UNIVERSITY.

Moghaddam, S., & Ester, M. (2010). Opinion Digger : An Unsupervised Opinion Miner from Unstructured Product Reviews. In *CIKM '10 Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 1825–1828).

Moghaddam, S., & Ester, M. (2012). Aspect-based Opinion Mining from Online Reviews. In *tutorial at SIGIR 2012*.

Moghaddam, S., & Ester, M. (2013). Opinion Mining in Online Reviews : Recent Trends. In *Tutorial at 22nd International World Wide Web Conference*.

Moghaddam, S., Jamali, M., & Ester, M. (2012). ETF : Extended Tensor Factorization Model for Personalizing Prediction of Review Helpfulness Categories and Subject Descriptors. In *Proceedings of the fifth ACM international conference on Web search and data mining series WSDM '12* (pp. 163–172).

Morinaga, S., Yamanishi, K., Tateishi, K., & Fukushima, T. (2002a). Mining product reputations on the Web. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 341–349). New York, NY, USA.

Morinaga, S., Yamanishi, K., Tateishi, K., & Fukushima, T. (2002b). Mining product reputations on the Web. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 341–349). New York, NY, USA: ACM. doi:10.1145/775047.775098

Mudambi, S. M., & Schuff, D. (2010). What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*, *34*(1), 185–200.

Mullen, T., & Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of EMNLP-2004.*

N. Au, D. Buhalis, and R. L. (2009). Complaints on the online environment the case of hong kong hotels. *Information and Communication Technologies in Tourism*, 73–85.

Na, J.-C., Thet, T. T., & Khoo, C. S. G. (2010). Comparing sentiment expression in movie reviews from four online genres. *Online Information Review*, *34*(2), 317–338. doi:10.1108/14684521011037016

Nagar, A., & Hahsler, M. (2012). Using Text and Data Mining Techniques to extract Stock Market Sentiment from Live News Streams. In *International Conference on Computer Technology and Science (ICCTS 2012)*. Singapore.

Ng, V., Dasgupta, S., & Arifin, S. M. N. (2006). Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews. In

Proceedings of COLING / ACL 2006 Main Conference Poster Sessions (COLING/ACL-2006). (pp. 611–618).

Ngo-Ye, T. L., & Sinha, A. P. (2012). Analyzing Online Review Helpfulness Using a Regressional ReliefF-Enhanced Text Mining Method. *ACM Transactions on Management Information Systems, 3*(2). doi:10.1145/2229156.2229158

Nightingale, F. (1858). *Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army: Founded Chiefly on the Experience of the Late War*.

O'Mahony, M. P., & Smyth, B. (2009). Learning to recommend helpful hotel reviews. In *Proceedings of the third ACM conference on Recommender systems - RecSys '09*. New York, New York, USA: ACM Press. doi:10.1145/1639714.1639774

Oelke, D., Hao, M., Rohrdantz, C., Keim, D. A., Dayal, U., Haug, L., & Janetzko, H. (2009). Visual Opinion Analysis of Customer Feedback Data. In *IEEE Symposium on Visual Analytics Science and Technology* (pp. 187–194).

Osimo, D., & Mureddu, F. (2012). Research Challenge on Opinion Mining and Sentiment Analysis. In *Using open data: policy modeling, citizen empowerment, data journalism*. Retrieved from http://www.w3.org/2012/06/pmod/opinionmining.pdf

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding Deceptive Opinion Spam by Any Stretch of the Imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011)* (pp. 309–319).

Paltoglou, G., & Thelwall, M. (2010). A study of Information Retrieval weighting schemes for sentiment analysis. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL-2010* (pp. 1386–1395).

Pan, S. J., Ni, X., Sun, J.-T., Yang, Q., & Chen, Z. (2010). Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th international conference on World wide web - WWW '10*. New York, New York, USA: ACM Press. doi:10.1145/1772690.1772767

Pana, Y., & Zhang, J. Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing, 87*(4), 598–612. doi:10.13189/aeb.2013.010101

Pang, B., & Lee, L. (2004). Seeing stars : Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of ACL-05, 43rd Meeting of the As- sociation for Computational Linguistics* (pp. 115–124). US.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval, 1*(2), 91–231. doi:10.1561/1500000001

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up ? Sentiment Classification using Machine Learning Techniques. In *Proc. EMNLP-02*.

Piao, S. S., Ananiadou, S., Tsuruoka, Y., Sasaki, Y., & Mcnaught, J. (2007). Mining Opinion Polarity Relations of Citations. *International Workshop on Computational Semantics (IWCS)*, 366–371.

Popescu, A.-M., & Etzioni, O. (2005). Extracting product features and opinions from reviews. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05* (pp. 339–346). Canada, Vancouver, B.C., Canada: Association for Computational Linguistics. doi:10.3115/1220575.1220618

Qiu, G., Liu, B., Bu, J., & Chen, C. (2011). Opinion Word Expansion and Target Extraction through Double Propagation. *Computational Linguistics*, *37*.

Raaijmakers, S., & Kraaij, W. (2008). A Shallow Approach to Subjectivity Classification. In *Proceedings of ICWSM-2008*.

Raghavan, P. (2009). *Extracting Opinions from Blog Comments: Analysis, Design and Applications*. Ohio State University.

Rohrdantz, C., Hao, M. C., Dayal, U., Haug, L.-E., & Keim, D. a. (2012a). Feature-Based Visual Sentiment Analysis of Text Document Streams. *ACM Transactions on Intelligent Systems and Technology*, *3*(2), 1–25. doi:10.1145/2089094.2089102

Rohrdantz, C., Hao, M. C., Dayal, U., Haug, L.-E., & Keim, D. a. (2012b). Feature-Based Visual Sentiment Analysis of Text Document Streams. *ACM Transactions on Intelligent Systems and Technology*, *3*(2), 1–25. doi:10.1145/2089094.2089102

Saggion, H., Funk, A., Street, P., & Sheffield, S. (2009). Extracting Opinions and Facts for Business Intelligence. *RNTI Journal*, (17), 119–146. Retrieved from http://gate.ac.uk/sale/rnti-09/final-version/Saggion-Funks-OM-09.pdf

Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H., & Jin, C. (2007). Red Opal : Product-Feature Scoring from Reviews. In *Proceedings of the 8th ACM conference on Electronic commerce* (pp. 182–191). San Diego, California, USA.

Seerat, B., & Azam, F. (2012). Opinion Mining : Issues and Challenges (A survey). *International Journal of Computer Applications*, *49*(9), 42–51.

Shamim, A., Balakrishnan, V., & Tahir, M. (2013). The Evaluation of Opinion Mining and Sentiment Analysis Systems using Evaluation Metrics. In *20th IBIMA Conference* (pp. 8–14). Kuala Lumpur, Malaysia.

Shamim, A., Balakrishnan, V., & Tahir, M. (2014). Evaluation of Radial and Hierarchical Opinion Visualization Techniques. In *International Conference on Advances in Engineering and Technology (ICAET'2014)* (pp. 442–446). Singapore.

Shariaty, S. (2011). Fine-Grained User Opinions Extraction Using Conditional Random Fields. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, ICDMW '11* (pp. 109–114). Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, ICDMW '11, pages 109–114, Washington, DC, USA: IEEE Computer Society.

Somasundaran, S., & Wiebe, J. (2009). Recognizing Stances in Online Debates. In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP (ACL-IJCNLP-2009)* (pp. 226–234).

Somasundaran, S., Wilson, T., Wiebe, J., & Stoyanov, V. (2007). QA with Attitude: Exploiting Opinion Type Analysis for Improving Question Answering in On-line Discussions and the News. In *International Conference on Weblogs and Social Media*. Retrieved from http://www.icwsm.org/papers/paper6.html

Somprasertsri Gamgarn, L. P. (2010). Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization. *Journal of Universal Computer Science*, *16*(6), 938–955.

Stylios, G., Christodoulakis, D., Besharat, J., Kotrotsos, I., Koumpouri, A., & Stamou, S. (2010). Public Opinion Mining for Governmental Decisions. *World Wide Web Internet And Web Information Systems*, *8*(2), 203–214.

Taboada, M., Brooke, J., & Voll, K. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267–307.

Tan, S., Wu, G., Tang, H., & Cheng, X. (2007). A novel scheme for domain-transfer problem in the context of sentiment analysis. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management - CIKM '07*. New York, New York, USA: ACM Press. doi:10.1145/1321440.1321590

Tang, J., Gao, H., Hu, X., & Liu, H. (2013). Context-aware review helpfulness rating prediction. *Proceedings of the 7th ACM Conference on Recommender Systems - RecSys '13*, 1–8. doi:10.1145/2507157.2507183

Tata, S., & Eugenio, B. Di. (2010). Generating Fine-Grained Reviews of Songs From Album Reviews. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010)* (pp. 1376–1385).

Titov, I., & Mcdonald, R. (n.d.). A Joint Model of Text and Aspect Ratings for Sentiment Summarization. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2008).*

Titov, I., & McDonald, R. (n.d.). Modeling online reviews with multi-grain topic models. In *Proceeding of the 17th international conference on World Wide Web - WWW '08* (p. 111). New York, New York, USA: ACM Press. doi:10.1145/1367497.1367513

Tsaparas, P., Ntoulas, A., & Terzi, E. (2011). Selecting a comprehensive set of reviews. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11* (p. 168). New York, New York, USA: ACM Press. doi:10.1145/2020408.2020440

Tsur, O., & Rappoport, A. (2009). R EV R ANK : a Fully Unsupervised Algorithm for Selecting the Most Helpful Book Reviews. In *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009,*. San Jose, California, USA,.

Turney, P. D. (2002). Thumbs Up or Thumbs Down ? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 02)* (pp. 417–424). Philadelphia.

Turney, P. D., & Littman, M. L. (2003). Measuring Praise and Criticism : Inference of Semantic Orientation from Association. *ACM Transactions on Information Systems (TOIS)*, *21*(4), 315–346.

Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, *30*(1), 123–127. doi:10.1016/j.tourman.2008.04.008

Walter, F. E., Battiston, S., & Schweitzer, F. (n.d.). Coping with Information Overload through Trust-Based Networks. *Managing Complexity: Insights, Concepts, Applications*, 273–300.

Wang, G., & Araki, K. (2007). OMS-J : An Opinion Mining System for Japanese Weblog Reviews Using a Combination of Supervised and Unsupervised Approaches. In *NAACL HLT Demonstration Program, Association for Computational Linguistics* (pp. 19–20). Rochester, New York, USA.

Wang, G., Xie, S., Liu, B., & Yu, P. S. (2012). Identify Online Store Review Spammers via Social Review Graph. *ACM Transactions on Intelligent Systems and Technology*, *3*(4), 1–21. doi:10.1145/2337542.2337546

Wanner, F., Rohrdantz, C., Mansmann, F., Oelke, D., & Keim, D. A. (2009). Visual Sentiment Analysis of RSS News Feeds Featuring the US Presidential Election in 2008. In *Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW2009)*. Sanibel Island, Florida, USA.

Wiebe, J. M., Bruce, R. F., & O'Hara, T. P. (1999). Development and Use of a Gold-Standard Data Set for Subjectivity Classifications. In *ACL '99 Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 246–253).

Wiebe, J., & Riloff, E. (2005). Creating Subjective and Objective Sentence Classifiers from Unannotated Texts. *Computational Linguistics and Intelligent Text Processing*, 486–497.

Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004). Learning Subjective Language. *Computational Linguistics, 30*(3), 277–308.

Wilson, T., Wiebe, J., & Hwa, R. (2004). Just how mad are you ? Finding strong and weak opinion clauses. In *Proc. 19th National Conference on Artificial Intelligence*.

Wu, Q., Tan, S., & Cheng, X. (2009). Graph Ranking for Sentiment Transfer. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers (ACL-IJCNLP-2009).*

Wu, Y., Wei, F., Liu, S., & Au, N. (2010). OpinionSeer : Interactive Visualization of Hotel Customer Feedback. *IEEE Transactions on Visualization and Computer Graphics, 16*(6), 1109–1118.

Wu, Y., Zhang, Q., Huang, X., & Wu, L. (2009). Phrase Dependency Parsing for Opinion Mining. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)* (pp. 1533–1541).

Xia, R., & Zong, C. (2011). A POS-based Ensemble Model for Cross-domain Sentiment Classification. In *Proceedings of the 5th International Joint Conference on Natural Language Processing.* (pp. 614–622).

Xu, K., Liao, S. S., Li, J., & Song, Y. (2011). Mining comparative opinions from customer reviews for Competitive Intelligence. *Decision Support Systems*, *50*(4), 743–754. doi:10.1016/j.dss.2010.08.021

Yang, H., Si, L., & Callan, Ja. (2006). Knowledge Transfer and Opinion Detection in the TREC2006 Blog Track. In *Proceedings of TREC*.

Yang, J.-Y., Kim, H.-J., & Lee, S.-G. (2010). Feature-based Product Review Summarization Utilizing User Score *. *Journal of Information Science and Engineering*, *26*, 1973–1990.

Yang, S. (2011). Extracting Comparative Entities and Predicates from Texts Using Comparative Type Classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (pp. 1636–1644).

Yu, H., & Hatzivassiloglou, V. (2008). Towards Answering Opinion Questions : Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. In *Proceedings of 8th Conference on Empirical Methods in Natural Language Processing (EMNLP-03)* (pp. 129–136).

Zhang, L. E. I. (2012). *Aspect and Entity Extraction from Opinion Documents*. University of Illinois at Chicago.

Zhang, L., & Liu, B. (2014). Aspect and Entity Extraction for Opinion Mining. In *Data Mining and Knowledge Discovery for Big Data: Methodologies, Challenges, and Opportunities*. Springer.

Zhang, Y., & Zhang, D. (2014). Automatically Predicting the Helpfulness of Online Reviews. In *IEEE 15th International Conference on Information Reuse and Integration (IRI)* (pp. 662–668).

Zhang, Z., & Varadarajan, B. (2006). Utility scoring of product reviews. In *Proceedings of the 15th ACM international conference on Information and knowledge management - CIKM '06*. New York, New York, USA: ACM Press. doi:10.1145/1183614.1183626

Zhao, W. X., Jiang, J., Yan, H., & Li, X. (2010). Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics.

Zhou, S., Chen, Q., & Wang, X. (2010). Active Deep Networks for Semi-Supervised Sentiment Classification. In *Proceedings of Coling 2010: Poster Volume. 2010.* (pp. 1515–1523).

Zhu, J., Wang, H., Tsou, B. K., & Zhu, M. (2009). Multi-aspect opinion polling from textual reviews. In *Proceeding of the 18th ACM conference on Information and knowledge management - CIKM '09*. New York, New York, USA: ACM Press. doi:10.1145/1645953.1646233

Zhuang, L., Jing, F., & Zhu, X.-Y. (2006). Movie Review Mining and Summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management - CIKM '06*. New York, New York, USA: ACM Press. doi:10.1145/1183614.1183625

List of Publications

1. Azra Shamim, Vimala Balakrishnan, Muhammad Tahir, Evaluation of Opinion Visualization Techniques in Information Visualization Journal, 2014

2. Azra Shamim, Vimala Balakrishnan, Muhammad Tahir, Muhammad Shiraz, Critical Product Features' Identification using an Opinion Analyzer in The Scientific World Journal, 2014

3. Azra Shamim,Vimala Balakrishnan, Muhammad Tahir, "Opinion Mining and Sentiment Analysis Systems: A Comparison of Design Considerations" published in 20th IBIMA conference, special edition, Kuala Lumpur, Malaysia, 25-26 March, 2013 , pp 1-7.

4. Azra Shamim,Vimala Balakrishnan, Muhammad Tahir, " The Evaluation of Opinion Mining and Sentiment Analysis Systems using Evaluation Metrics " published in 20th IBIMA conference, special edition, Kuala Lumpur, Malaysia, 25-26 March, 2013 , pp 8-14.

5. Azra Shamim, Vimala Balakrishnan, Muhammad Tahir, Evaluation of Radial and Hierarchical Opinion Visualization Techniques, In International Conference on Advances in Engineering and Technology (ICAET'2014), March 29-30, 2014, Singapore.